BEST AVAILABLE COPY

Access DB# 40715

CRFE

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Natalie Davis Examiner #: 78462 Date: 4-23-01

Art Unit: 1642 Phone Number 308-6410 Serial Number: 09/721183

Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

7E12 CM1 4C01

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: 11-23-99

For Sequence Searches Only *Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Please search SEQ ID NO: 1, 2, 3, 4, 5, 18 & 20

RECEIVED APR 23 2001 (STIC)

Point of Contact
Beverly Shears
Technical Info. Specialist
CM1 12C14 Tel: 308-4994

STAFF USE ONLY	Type of Search	Vendors and cost where applicable
Searcher: Beverly @4994	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: 05-01-01	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: 10	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: 20	Other _____	Other (specify) CGN

PTO-1590 (1-2000)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 08:52:06 ; Search time 22667.1 Seconds
(without alignments)
292.858 Million cell updates/sec

Title: US-09-721-183-1
Perfect score: 1078
Sequence: 1 ggatgatacaagagccaaga.....tggtgttcctttgaatcgtg 1078

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pri:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

This Page Blank (uspto)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	1065.4	98.8	170388	71	AC027750	AC027750 Homo sapi
c	2	1065.4	98.8	206319	69	AC025097	AC025097 Homo sapi
	3	55.4	5.1	125977	83	CNS05TC5	AL355053 Homo sapi
	4	51.6	4.8	133488	86	AC007963	AC007963 Homo sapi
	5	51.6	4.8	174713	76	AC079083	AC079083 Homo sapi
	6	51	4.7	179899	64	AC015756	AC015756 Homo sapi
	7	50	4.6	151505	67	AC021875	AC021875 Homo sapi
	8	50	4.6	159066	86	AC006417	AC006417 Homo sapi
c	9	49.4	4.6	165567	75	AC073886	AC073886 Homo sapi
c	10	48.6	4.5	896	53	CNS06FEG	AL396302 T3 end of
	11	48.4	4.5	7218	10	I66494	I66494 Sequence 14

c	12	48	4.5	182336	77	AC079785	AC079785 Homo sapi
	13	47.8	4.4	964	54	CNS07E8R	AL441457 T7 end of
	14	47.8	4.4	147155	62	AC011061	AC011061 Homo sapi
	15	47.8	4.4	182444	61	AC009451	AC009451 Homo sapi
c	16	47.8	4.4	199534	70	AC025959	AC025959 Homo sapi
c	17	47.6	4.4	61869	70	AC026007	AC026007 Homo sapi
	18	47.6	4.4	132205	61	AC009271	AC009271 Homo sapi
	19	47.6	4.4	165928	71	AC027584	AC027584 Homo sapi
c	20	47.6	4.4	172371	68	AC023468	AC023468 Homo sapi
	21	47	4.4	173328	66	AC019162	AC019162 Homo sapi
c	22	46.8	4.3	177212	70	AC025733	AC025733 Homo sapi
c	23	46.6	4.3	973	53	CNS071LE	AL425064 clone BA0
c	24	46.6	4.3	110991	78	AF165144	AF165144 Homo sapi
c	25	46.6	4.3	173943	68	AC022717	AC022717 Homo sapi
	26	46.6	4.3	183068	71	AC027595	AC027595 Homo sapi
c	27	46	4.3	175270	67	AC021465	AC021465 Homo sapi
c	28	46	4.3	211039	74	AC068555	AC068555 Homo sapi
c	29	45.8	4.2	886	53	CNS075DF	AL429961 clone BA0
c	30	45.2	4.2	1007	53	CNS06X9S	AL419462 T3 end of
	31	45.2	4.2	253305	6	PFMAL3P7	AL034559 Plasmodiu
	32	44.8	4.2	110000	84	PFMAL4P1_0	AL034557 Plasmodiu
	33	44.8	4.2	111395	71	AC034206	AC034206 Homo sapi
	34	44.8	4.2	143436	85	AC004456	AC004456 Homo sapi
	35	44.8	4.2	170805	89	AL133321	AL133321 Human DNA
c	36	44.4	4.1	924	54	CNS07A5L	AL436159 T7 end of
	37	44.4	4.1	158700	71	AC027643	AC027643 Homo sapi
c	38	44.2	4.1	10213	6	PFU27338	U27338 Plasmodium
	39	44.2	4.1	190239	61	AC009522	AC009522 Homo sapi
	40	44.2	4.1	231952	74	AC068641	AC068641 Homo sapi
c	41	44	4.1	1381	4	AE002979	AE002979 Drosophil
	42	44	4.1	146250	76	AC074240	AC074240 Homo sapi
c	43	43.8	4.1	895	53	CNS01KDS	AL148209 Anopheles
c	44	43.8	4.1	895	53	CNS06FSV	AL396821 T7 end of
	45	43.8	4.1	110388	78	AF206725	AF206725 Homo sapi

ALIGNMENTS

```
RESULT   1
AC027750/c
LOCUS       AC027750   170388 bp    DNA              HTG       02-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-607P23 map 4, WORKING DRAFT
            SEQUENCE, 28 unordered pieces.
ACCESSION   AC027750
VERSION     AC027750.3  GI:7677886
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170388)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-607P23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170388)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 2, 2000 this sequence version replaced gi:7387407.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L8208
                Center clone name: 607_P_23
            ------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156218 bases at least Q40
                Consensus quality: 163157 bases at least Q30
                Consensus quality: 165910 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 167688; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1255: contig of 1255 bp in length
            *     1256 1355: gap of      100 bp
            *     1356     2376: contig of 1021 bp in length
            *     2377 2476: gap of      100 bp
            *     2477     3971: contig of 1495 bp in length
            *     3972 4071: gap of      100 bp
            *     4072     5644: contig of 1573 bp in length
            *     5645 5744: gap of      100 bp
            *     5745     7371: contig of 1627 bp in length
            *     7372 7471: gap of      100 bp
            *     7472     9665: contig of 2194 bp in length
            *     9666 9765: gap of      100 bp
            *     9766    11555: contig of 1790 bp in length
            *    11556 11655: gap of      100 bp
            *    11656    13660: contig of 2005 bp in length
            *    13661 13760: gap of      100 bp
            *    13761    18045: contig of 4285 bp in length
            *    18046 18145: gap of      100 bp
            *    18146    21359: contig of 3214 bp in length
            *    21360 21459: gap of      100 bp
            *    21460    24888: contig of 3429 bp in length
            *    24889 24988: gap of      100 bp
            *    24989    28915: contig of 3927 bp in length
            *    28916 29015: gap of      100 bp
            *    29016    32863: contig of 3848 bp in length
            *    32864 32963: gap of      100 bp
            *    32964    37173: contig of 4210 bp in length
            *    37174 37273: gap of      100 bp
            *    37274    42567: contig of 5294 bp in length
            *    42568 42667: gap of      100 bp
            *    42668    47610: contig of 4943 bp in length
            *    47611 47710: gap of      100 bp
            *    47711    53661: contig of 5951 bp in length
            *    53662 53761: gap of      100 bp
```

```
          *    53762    59606: contig of 5845 bp in length
          *    59607 59706: gap of      100 bp
          *    59707    67728: contig of 8022 bp in length
          *    67729 67828: gap of      100 bp
          *    67829    73763: contig of 5935 bp in length
          *    73764 73863: gap of      100 bp
          *    73864    83669: contig of 9806 bp in length
          *    83670 83769: gap of      100 bp
          *    83770    92316: contig of 8547 bp in length
          *    92317 92416: gap of      100 bp
          *    92417   104093: contig of 11677 bp in length
          *   104094 104193: gap of      100 bp
          *   104194   116583: contig of 12390 bp in length
          *   116584 116683: gap of      100 bp
          *   116684   128499: contig of 11816 bp in length
          *   128500 128599: gap of      100 bp
          *   128600   141281: contig of 12682 bp in length
          *   141282 141381: gap of      100 bp
          *   141382   154335: contig of 12954 bp in length
          *   154336 154435: gap of      100 bp
          *   154436   170388: contig of 15953 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-607P23"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1255
                     /note="assembly_fragment"
     misc_feature    1356. .2376
                     /note="assembly_fragment"
     misc_feature    2477. .3971
                     /note="assembly_fragment"
     misc_feature    4072. .5644
                     /note="assembly_fragment"
     misc_feature    5745. .7371
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    7472. .9665
                     /note="assembly_fragment"
     misc_feature    9766. .11555
                     /note="assembly_fragment"
     misc_feature    11656. .13660
                     /note="assembly_fragment"
     misc_feature    13761. .18045
                     /note="assembly_fragment"
     misc_feature    18146. .21359
                     /note="assembly_fragment"
     misc_feature    21460. .24888
                     /note="assembly_fragment"
     misc_feature    24989. .28915
                     /note="assembly_fragment"
     misc_feature    29016. .32863
                     /note="assembly_fragment"
     misc_feature    32964. .37173
                     /note="assembly_fragment"
     misc_feature    37274. .42567
                     /note="assembly_fragment"
     misc_feature    42668. .47610
                     /note="assembly_fragment"
     misc_feature    47711. .53661
                     /note="assembly_fragment"
     misc_feature    53762. .59606
                     /note="assembly_fragment"
     misc_feature    59707. .67728
                     /note="assembly_fragment"
     misc_feature    67829. .73763
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    73864. .83669
                     /note="assembly_fragment"
     misc_feature    83770. .92316
                     /note="assembly_fragment"
     misc_feature    92417. .104093
                     /note="assembly_fragment"
     misc_feature    104194. .116583
                     /note="assembly_fragment"
     misc_feature    116684. .128499
                     /note="assembly_fragment"
     misc_feature    128600. .141281
                     /note="assembly_fragment"
     misc_feature    141382. .154335
                     /note="assembly_fragment"
     misc_feature    154436. .170388
                     /note="assembly_fragment"
BASE COUNT    46971 a  36488 c  37534 g  46693 t   2702 others
ORIGIN
```

Query Match 98.8%; Score 1065.4; DB 71; Length 170388;
Best Local Similarity 99.8%; Pred. No. 9.1e-224;
Matches 1077; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy       1 ggatgatacaagagccaagaagggacatttgagttgtgtcgcttagataggaaagggatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141119 GGATGATACAAGAGCCAAGAAGGGACATTTGAGTTGTGTCGCTTAGATAGGAAAGGGATC 141060

Qy      61 cagggaaaatcaacagtaagtgaggatgagcagcgtctcttggttttcattgaggataga 120
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 141059 CAGGGAAAATCAACAGTAAGTGAGGATGAGCAGTGTCTCTTGGTTTTCATTGAGGATAGA 141000

Qy     121 gtaagagattgagtttagattgcaacagaaggaattagtttagataccaggaagaacttc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140999 GTAAGAGATTGAGTTTAGATTGCAACAGAAGGAATTAGTTTAGATACCAGGAAGAACTTC 140940

Qy     181 ctagcctgaagatttgtcatagtgtctgctttctagatatctgggaaagatttgataata 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140939 CTAGCCTGAAGATTTGTCATAGTGTCTGCTTTCTAGATATCTGGGAAAGATTTGATAATA 140880

Qy     241 gttgtttgtgaatagaaaggaggatatgatgtttttattggccattttgcgggactcttc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140879 GTTGTTTGTGAATAGAAAGGAGGATATGATGTTTTTATTGGCCATTTTGCGGGACTCTTC 140820

Qy     301 gacttcttgctgctgtctcttgaggatacattccaattccatcctggcgagatccaagtg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140819 GACTTCTTGCTGCTGTCTCTTGAGGATACATTCCAATTCCATCCTGGCGAGATCCAAGTG 140760

Qy     361 cttacgtactgtctccttagctgccttagagtaaacgatcatcagttcaatggaccaaaa 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140759 CTTACGTACTGTCTCCTTAGCTGCCTTAGAGTAAACGATCATCAGTTCAATGGACCAAAA 140700

Qy     421 tcaccttcagccatgtggtttcttcatcatcatggatttcttttggttgacaaacattct 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140699 TCACCTTCAGCCATGTGGTTTCTTCATCATCATGGATTTCTTTTGGTTGACAAACATTCT 140640

Qy     481 ggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtgaaattagttttg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140639 GGCTCTCAGATGCAAAAAGTCACACTGGGAAATGAACTGTAAGTGGTGAAATTAGTTTTG 140580

Qy     541 gtatttaatttaaaactacattt-tagtttttctcttctcttctatgttgcaatgaatgt 599
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 140579 GTATTTAATTTAAAACTACATTTATAGTTTTTCTCTTCTCTTCTATGTTGCAATGAATGT 140520

Qy     600 aaagtatttgggatccagtgcttataaacctttccttcctttgtgcacagaatgtaacta 659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140519 AAAGTATTTGGGATCCAGTGCTTATAAACCTTTCCTTCCTTTGTGCACAGAATGTAACTA 140460

Qy     660 gcaagcccattagcacccagataattctatcatgttagtttcccatcctggaaaatcttt 719
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140459 GCAAGCCCATTAGCACCCAGATAATTCTATCATGTTAGTTTCCCATCCTGGAAAATCTTT 140400

Qy     720 gtacagtgggaagttccccgatgtgtttttctttcttaggtgaagggttggctatatcac 779
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140399 GTACAGTGGGAAGTTCCCCGATGTGTTTTTCTTTCTTAGGTGAAGGGTTGGCTATATCAC 140340

Qy    780 tttattgaattttgcattccttagacttttaaaatatactaatgtattctagtcttactc 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140339 TTTATTGAATTTTGCATTCCTTAGACTTTTAAAATATACTAATGTATTCTAGTCTTACTC 140280

Qy    840 taaagacctttgatgttaaaggaatccttcatttatttcatattccctatctcatagggc 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140279 TAAAGACCTTTGATGTTAAAGGAATCCTTCATTTATTTCATATTCCCTATCTCATAGGGC 140220

Qy    900 cacaattattttaatacagagatgattttcaaaatattttaacaactggtacaggacaga 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140219 CACAATTATTTTAATACAGAGATGATTTTCAAAATATTTTAACAACTGGTACAGGACAGA 140160

Qy    960 tgccagccactcagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatg 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140159 TGCCAGCCACTCAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATG 140100

Qy   1020 cctattggctgaacattatgctactttcagatattaaaatggtgttcctttgaatcgtg 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140099 CCTATTGGCTGAACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGTG 140041
```

RESULT 2
AC025097/c
LOCUS AC025097 206319 bp DNA HTG 02-FEB-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-285E4, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC025097
VERSION AC025097.34 GI:12656669
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 206319)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 206319)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Feb 2, 2001 this sequence version replaced gi:12621181.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HAMD
Center clone name: RP11-285E4
---------- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 7% of reads
Chemistry: Dye-terminator Big Dye: 93% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 206515 bases at least Q40
Consensus quality: 207810 bases at least Q30
Consensus quality: 208593 bases at least Q20
Estimated insert size: 204397; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 10.5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 54724: contig of 54724 bp in length
* 54725 54824: gap of unknown length
* 54825 110676: contig of 55852 bp in length
* 110677 110776: gap of unknown length
* 110777 155330: contig of 44554 bp in length
* 155331 155430: gap of unknown length
* 155431 205162: contig of 49732 bp in length
* 205163 205262: gap of unknown length
* 205263 206319: contig of 1057 bp in length.
FEATURES Location/Qualifiers
source 1. .206319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-285E4"
BASE COUNT 60783 a 42912 c 42453 g 59760 t 411 others
ORIGIN

Query Match 98.8%; Score 1065.4; DB 69; Length 206319;
Best Local Similarity 99.8%; Pred. No. 9e-224;
Matches 1077; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy      1 ggatgatacaagagccaagaagggacatttgagttgtgtcgcttagataggaaagggatc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139445 GGATGATACAAGAGCCAAGAAGGGACATTTGAGTTGTGTCGCTTAGATAGGAAAGGGATC 139386

Qy     61 cagggaaaatcaacagtaagtgaggatgagcagcgtctcttggttttcattgaggataga 120
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 139385 CAGGGAAAATCAACAGTAAGTGAGGATGAGCAGTGTCTCTTGGTTTTCATTGAGGATAGA 139326
```

```
Qy    121 gtaagagattgagtttagattgcaacagaaggaattagtttagataccaggaagaacttc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139325 GTAAGAGATTGAGTTTAGATTGCAACAGAAGGAATTAGTTTAGATACCAGGAAGAACTTC 139266

Qy    181 ctagcctgaagatttgtcatagtgtctgctttctagatatctgggaaagatttgataata 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139265 CTAGCCTGAAGATTTGTCATAGTGTCTGCTTTCTAGATATCTGGGAAAGATTTGATAATA 139206

Qy    241 gttgtttgtgaatagaaaggaggatatgatgtttttattggccattttgcgggactcttc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139205 GTTGTTTGTGAATAGAAAGGAGGATATGATGTTTTTATTGGCCATTTTGCGGGACTCTTC 139146

Qy    301 gacttcttgctgctgtctcttgaggatacattccaattccatcctggcgagatccaagtg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139145 GACTTCTTGCTGCTGTCTCTTGAGGATACATTCCAATTCCATCCTGGCGAGATCCAAGTG 139086

Qy    361 cttacgtactgtctccttagctgccttagagtaaacgatcatcagttcaatggaccaaaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139085 CTTACGTACTGTCTCCTTAGCTGCCTTAGAGTAAACGATCATCAGTTCAATGGACCAAAA 139026

Qy    421 tcaccttcagccatgtggtttcttcatcatcatggatttcttttggttgacaaacattct 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 139025 TCACCTTCAGCCATGTGGTTTCTTCATCATCATGGATTTCTTTTGGTTGACAAACATTCT 138966

Qy    481 ggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtgaaattagttttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138965 GGCTCTCAGATGCAAAAAGTCACACTGGGAAATGAACTGTAAGTGGTGAAATTAGTTTTG 138906

Qy    541 gtatttaatttaaaactacattt-tagtttttctcttctcttctatgttgcaatgaatgt 599
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 138905 GTATTTAATTTAAAACTACATTTATAGTTTTTCTCTTCTCTTCTATGTTGCAATGAATGT 138846

Qy    600 aaagtatttgggatccagtgcttataaacctttccttcctttgtgcacagaatgtaacta 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138845 AAAGTATTTGGGATCCAGTGCTTATAAACCTTTCCTTCCTTTGTGCACAGAATGTAACTA 138786

Qy    660 gcaagcccattagcacccagataattctatcatgttagtttcccatcctggaaaatcttt 719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138785 GCAAGCCCATTAGCACCCAGATAATTCTATCATGTTAGTTTCCCATCCTGGAAAATCTTT 138726

Qy    720 gtacagtgggaagttccccgatgtgtttttctttcttaggtgaagggttggctatatcac 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138725 GTACAGTGGGAAGTTCCCCGATGTGTTTTTCTTTCTTAGGTGAAGGGTTGGCTATATCAC 138666

Qy    780 tttattgaattttgcattccttagacttttaaaatatactaatgtattctagtcttactc 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138665 TTTATTGAATTTTGCATTCCTTAGACTTTTAAAATATACTAATGTATTCTAGTCTTACTC 138606

Qy    840 taaagacctttgatgttaaaggaatccttcatttatttcatattccctatctcatagggc 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138605 TAAAGACCTTTGATGTTAAAGGAATCCTTCATTTATTTCATATTCCCTATCTCATAGGGC 138546

Qy    900 cacaattattttaatacagagatgattttcaaaatattttaacaactggtacaggacaga 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138545 CACAATTATTTTAATACAGAGATGATTTTCAAAATATTTTAACAACTGGTACAGGACAGA 138486

Qy    960 tgccagccactcagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatg 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138485 TGCCAGCCACTCAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATG 138426

Qy   1020 cctattggctgaacattatgctactttcagatattaaaatggtgttcctttgaatcgtg 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138425 CCTATTGGCTGAACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGTG 138367
```

```
RESULT   3
CNS05TC5
LOCUS       CNS05TC5   125977 bp    DNA             HTG       08-DEC-2000
DEFINITION  Homo sapiens chromosome 14 clone C-2305P23, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL355053
VERSION     AL355053.4  GI:11611171
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125977)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:9453785.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continue. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage, etc. . . even if efforts are made to eliminate these
            contaminating sequences. The following BAC sequence is oriented
            from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-269C4 (AC=AL133372)
            Downstream BAC (overlapping the SP6 end) : R-419C10 --------------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 12.39x in Q20 bases; sum-of-contigs
            ---------------
            ------------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :
            10 - 19   :
            20 - 29   :      12
            30 - 39   :      69
            40 - 49   :    1319
            50 - 59   :    3363
            60 - 69   :    3948
            70 - 79   :    8516
            80 - 89   :   25131
            90 - 99   :   83619
            ------------------------
            Percentage of bases with a quality value >= 40 : 99 %.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES                Location/Qualifiers
     source             1. .125977
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="14"
                        /clone="C-2305P23"
                        /clone_lib="CalTech-D"
BASE COUNT     36262 a  23215 c  24088 g  42412 t
ORIGIN
```

```
  Query Match             5.1%;  Score 55.4;  DB 83;  Length 125977;
  Best Local Similarity   59.3%;  Pred. No. 0.025;
  Matches  112;  Conservative    0;  Mismatches   76;  Indels    1;  Gaps    1;

Qy    731 agttccccgatgtgtttttctttcttaggtgaagggttggctatatcactttattgaatt 790
          | | || |  |||  |      ||| | |       | | |  ||||||| |||| |||
Db  44246 ATTACCTCTTTGTCCTAACAAGTCTCAAATTCCTTCCTTGATTGATCACTTGATTGGATT 44305

Qy    791 ttgcattccttagacttttaaaatatactaatgtattctagtcttactctaaagaccttt 850
          ||  ||||| |   || |||| |||  | ||| | || |     |||   ||||||  ||
```

```
Db  44306 TTTAATTCCGTTAGCTCTTAAGATAGGCAAATCTCTTTTTTATCTACATGAAAGACTATT 44365

Qy    851 gatgttaaaggaatccttcatttatttcatattccctatctcatagggccacaattattt 910
          || ||| | | ||| || || ||   | || ||  || |||| || ||| ||| | |||
Db  44366 GAAGTTGATGAAATACTACACTTTCATAATCTT-TCTTTCTCTTAAGGCTACATTAATTC 44424

Qy    911 taatacaga 919
          | ||| | |
Db  44425 TTATAGATA 44433


RESULT   4
AC007963
LOCUS       AC007963   133488 bp    DNA             PRI       13-DEC-2000
DEFINITION  Homo sapiens chromosome 17, clone CTD-2010G8, complete sequence.
ACCESSION   AC007963
VERSION     AC007963.7  GI:11597104
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133488)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2010G8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 133488)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 133488)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 7, 2000 this sequence version replaced gi:11527473.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L889
                Center clone name: 2010_G_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .133488
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2010G8"
                     /clone_lib="CITD Human BAC"
     repeat_region   complement(4. .321)
                     /rpt_family="L1MDa"
     repeat_region   complement(342. .494)
                     /rpt_family="AluSq"
     repeat_region   complement(550. .637)
                     /rpt_family="MER66A"
     repeat_region   complement(638. .775)
                     /rpt_family="L1MDa"
     repeat_region   complement(965. .1444)
                     /rpt_family="L1MDa"
     repeat_region   complement(1445. .1742)
                     /rpt_family="AluSg"
     repeat_region   complement(1743. .1971)
                     /rpt_family="L1MDa"
     repeat_region   complement(1972. .2162)
                     /rpt_family="L1MB8"
     repeat_region   complement(2193. .2437)
                     /rpt_family="L1MDa"
     repeat_region   complement(2448. .2599)
                     /rpt_family="L1MDa"
     repeat_region   complement(2600. .2899)
                     /rpt_family="AluSx"
     repeat_region   complement(2900. .4288)
                     /rpt_family="L1MDa"
     repeat_region   complement(4289. .4585)
                     /rpt_family="AluSp"
     repeat_region   complement(4586. .4663)
                     /rpt_family="L1MDa"
     repeat_region   complement(4664. .4967)
                     /rpt_family="AluY"
     repeat_region   complement(4968. .6064)
                     /rpt_family="L1MDa"
     repeat_region   complement(6312. .6807)
                     /rpt_family="L1MC4a"
     repeat_region   6957. .7081
                     /rpt_family="L2"
     repeat_region   complement(8359. .8534)
                     /rpt_family="L1MC4a"
     repeat_region   complement(9165. .9287)
                     /rpt_family="FLAM_A"
     repeat_region   9298. .9341
                     /rpt_family="AT_rich"
     repeat_region   complement(9458. .10092)
                     /rpt_family="L2"
     repeat_region   complement(10222. .10648)
                     /rpt_family="HAL1b"
     repeat_region   complement(10649. .10896)
                     /rpt_family="AluY"
     repeat_region   complement(10897. .10940)
                     /rpt_family="HAL1b"
     repeat_region   complement(11179. .11474)
                     /rpt_family="AluY"
     repeat_region   11667. .11951
                     /rpt_family="AluJo"
     repeat_region   complement(12106. .12361)
```

```
                /rpt_family="L1MC/D"
   repeat_region   12362. .12467
                /rpt_family="FLAM_A"
   repeat_region   complement(12468. .12619)
                /rpt_family="L1MC/D"
   repeat_region   complement(12851. .13158)
                /rpt_family="AluSp"
   repeat_region   13181. .13279
                /rpt_family="L1PA5"
   repeat_region   14090. .14223
                /rpt_family="(TTTC)n"
   repeat_region   complement(14226. .14508)
                /rpt_family="AluSq"
   repeat_region   complement(14640. .14812)
                /rpt_family="L2"
   repeat_region   14875. .15121
                /rpt_family="MIR"
   repeat_region   16351. .16433
                /rpt_family="L2"
   repeat_region   16434. .16962
                /rpt_family="MLT1F1"
   repeat_region   16963. .17269
                /rpt_family="L2"
   repeat_region   complement(17700. .17919)
                /rpt_family="AluJb"
   repeat_region   complement(17935. .18553)
                /rpt_family="L1MB1"
   repeat_region   19255. .19291
                /rpt_family="AT_rich"
   repeat_region   19857. .19900
                /rpt_family="AT_rich"
   repeat_region   20194. .20508
                /rpt_family="AluY"
   repeat_region   complement(20524. .20825)
                /rpt_family="AluJo"
   repeat_region   21382. .21701
                /rpt_family="AluJb"
   repeat_region   complement(21748. .21861)
                /rpt_family="MER1A"
   repeat_region   complement(21862. .22173)
                /rpt_family="AluY"
   repeat_region   complement(22174. .22534)
                /rpt_family="MER1A"
   repeat_region   22717. .22925
                /rpt_family="AluJo"
   repeat_region   23399. .23533
                /rpt_family="AluJo"
   repeat_region   complement(24103. .24405)
                /rpt_family="AluSx"
   repeat_region   complement(24783. .24835)
                /rpt_family="L2"
   repeat_region   25012. .25306
                /rpt_family="AluY"
   repeat_region   complement(25416. .25697)
                /rpt_family="AluJo"
   repeat_region   complement(25727. .26021)
                /rpt_family="AluSx"
   repeat_region   complement(26133. .26340)
                /rpt_family="L1ME"
   repeat_region   complement(26401. .26590)
                /rpt_family="L1ME"
   repeat_region   complement(28133. .28971)
                /rpt_family="MER21A"
   repeat_region   29011. .29060
                /rpt_family="(TCTA)n"
   repeat_region   complement(29786. .30084)
                /rpt_family="AluJb"

 Query Match             4.8%;  Score 51.6;  DB 86;  Length 133488;
 Best Local Similarity   60.8%;  Pred. No. 0.17;
 Matches  101;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy     769 ggctatatcactttattgaattttgcattccttagacttttaaaatatactaatgtattc 828
           ||| ||||||| | |||  | ||| || | |||   |  |||||||||  ||| | |||
Db  127897 GGCCATATCACCTGATTAGAGTTTACACTTCTTTAGCACTTAAAATATATAAATATTTTC 127956

Qy     829 tagtcttactctaaagacctttgatgttaaaggaatccttc-atttatttcatattccct 887
             | | |||| ||||| ||||||  | | || |  |   |    |  || ||| ||| |  ||
Db  127957 CCATTTAACTCCAAAGAACTTTGACTTGATAGAATACTGCCTGGTTTTTTTATACTTTCT 128016

Qy     888 atctcatagggccacaattattttaatacagagatgattttcaaaa 933
           |||| ||  ||   |  | |||||| | |   |   || |   ||
Db  128017 GTCTCTTAAAGCAGTATCTTTTTTAAAAAAATTAATTTTATTTTAA 128062


RESULT   5
AC079083
LOCUS       AC079083   174713 bp    DNA             HTG       10-NOV-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-41H3 map 4, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC079083
VERSION     AC079083.4  GI:11136771
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-41H3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkly,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 10, 2000 this sequence version replaced gi:10440661.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10450
                Center clone name: 41_H_3
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 54% of reads
                Sequencing vector: Plasmid; n/a; 46% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 170519 bases at least Q40
                Consensus quality: 172315 bases at least Q30
                Consensus quality: 173210 bases at least Q20
                Insert size: 170000; agarose-fp
```

```
            Insert size: 173813; sum-of-contigs
            Quality coverage: 7.6 in Q20 bases; agarose-fp
            Quality coverage: 7.5 in Q20 ba.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    35971: contig of 35971 bp in length
          *    35972 36071: gap of      100 bp
          *    36072    36859: contig of 788 bp in length
          *    36860 36959: gap of      100 bp
          *    36960    38300: contig of 1341 bp in length
          *    38301 38400: gap of      100 bp
          *    38401    40427: contig of 2027 bp in length
          *    40428 40527: gap of      100 bp
          *    40528    42753: contig of 2226 bp in length
          *    42754 42853: gap of      100 bp
          *    42854   131670: contig of 88817 bp in length
          *   131671 131770: gap of      100 bp
          *   131771   140138: contig of 8368 bp in length
          *   140139 140238: gap of      100 bp
          *   140239   152462: contig of 12224 bp in length
          *   152463 152562: gap of      100 bp
          *   152563   169648: contig of 17086 bp in length
          *   169649 169748: gap of      100 bp
          *   169749   174713: contig of 4965 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174713
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-41H3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .35971
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    36072. .36859
                     /note="assembly_fragment"
     misc_feature    36960. .38300
                     /note="assembly_fragment"
     misc_feature    38401. .40427
                     /note="assembly_fragment"
     misc_feature    40528. .42753
                     /note="assembly_fragment"
     misc_feature    42854. .131670
                     /note="assembly_fragment"
     misc_feature    131771. .140138
                     /note="assembly_fragment"
     misc_feature    140239. .152462
                     /note="assembly_fragment"
     misc_feature    152563. .169648
                     /note="assembly_fragment"
     misc_feature    169749. .174713
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    47490 a  35870 c  37148 g  53305 t    900 others
ORIGIN


  Query Match             4.8%;  Score 51.6;  DB 76;  Length 174713;
  Best Local Similarity   60.8%;  Pred. No. 0.17;
  Matches  101;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy    769 ggctatatcactttattgaattttgcattccttagactttaaaatatactaatgtattc 828
          ||| ||||||| | |||  | ||| || | |||   |  ||||||||||  ||| | |||
Db  45109 GGCCATATCACCTGATTAGAGTTTACACTTCTTTAGCACTTAAAATATATAAATATTTTC 45168

Qy    829 tagtcttactctaaagacctttgatgttaaaggaatccttc-atttatttcatattccct 887
          | | |||| ||||| ||||||  | | || |  |   |   || ||| ||| |  ||
Db  45169 CCATTTAACTCCAAAGAACTTTGACTTGATAGAATACTGCCTGGTTTTTTTATACTTTCT 45228

Qy    888 atctcatagggccacaattattttaatacagagatgattttcaaaa 933
          |||| ||  ||   |  | |||||| | |   |   || |   ||
Db  45229 GTCTCTTAAAGCAGTATCTTTTTTAAAAAAATTAATTTTATTTTAA 45274


RESULT   6
AC015756
LOCUS       AC015756   179899 bp    DNA             HTG       29-NOV-2000
DEFINITION  Homo sapiens clone RP11-1L12, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC015756
VERSION     AC015756.5  GI:11415188
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179899)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-1L12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179899)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 29, 2000 this sequence version replaced gi:7341717.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1478
                Center clone name: 1_L_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 50% of reads
                Sequencing vector: Plasmid; n/a; 50% of reads
                Chemistry: Dye-primer-amersham; 2% of reads
                Chemistry: Dye-terminator Big Dye; 98% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 179476 bases at least Q40
                Consensus quality: 179732 bases at least Q30
                Consensus quality: 179784 bases at least Q20
                Insert size: 183000; agarose-fp
                Insert size: 179799; sum-of-contigs
                Quality coverage: 10.7 in Q20 bases;.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
```

```
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   120417: contig of 120417 bp in length
            *  120418 120517: gap of      100 bp
            *  120518   179899: contig of 59382 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179899
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-1L12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .120417
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    120518. .179899
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    56015 a  36742 c  34172 g  52870 t    100 others
ORIGIN


  Query Match             4.7%;  Score 51;  DB 64;  Length 179899;
  Best Local Similarity   74.8%;  Pred. No. 0.23;
  Matches   77;  Conservative    0;  Mismatches   25;  Indels    1;  Gaps    1;

Qy    771 ctatatcactttattgaattttgcattccttagacttttaaaatatactaatgtattct- 829
          |||| |||  |||| || ||||| || |  |||| | | |||||||   || || || |
Db 144456 CTATCTCATGTTATAGACTTTTGAATGCAATAGAATCTAAAAATATGTGAACGTCTTTTC 144515

Qy    830 agtcttactctaaagacctttgatgttaaaggaatccttcatt 872
          ||||| |||||||||| |||||   |||| |||||||| ||||
Db 144516 AGTCTAACTCTAAAGATCTTTGTGCTTAATGGAATCCTGCATT 144558


RESULT   7
AC021875
LOCUS       AC021875   151505 bp    DNA             HTG       07-JAN-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-463I23, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC021875
VERSION     AC021875.16  GI:12025587
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151505)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151505)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 4, 2001 this sequence version replaced gi:11128302.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMZG
                Center clone name: RP11-463I23
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 14% of reads
                Chemistry: Dye-terminator Big Dye: 86% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 134190 bases at least Q40
                Consensus quality: 143515 bases at least Q30
                Consensus quality: 147123 bases at least Q20
                Estimated insert size: 148215; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    25621: contig of 25621 bp in length
            *   25622    25721: gap of unknown length
            *   25722    50610: contig of 24889 bp in length
            *   50611    50710: gap of unknown length
            *   50711    78610: contig of 27900 bp in length
            *   78611    78710: gap of unknown length
            *   78711    91238: contig of 12528 bp in length
            *   91239    91338: gap of unknown length
            *   91339   101762: contig of 10424 bp in length
            *  101763   101862: gap of unknown length
            *  101863   111900: contig of 10038 bp in length
            *  111901   112000: gap of unknown length
            *  112001   122047: contig of 10047 bp in length
            *  122048   122147: gap of unknown length
            *  122148   127963: contig of 5816 bp in length
            *  127964   128063: gap of unknown length
            *  128064   136011: contig of 7948 bp in length
```

```
            *   136012   136111: gap of unknown length
            *   136112   140698: contig of 4587 bp in length
            *   140699   140798: gap of unknown length
            *   140799   144700: contig of 3902 bp in length
            *   144701   144800: gap of unknown length
            *   144801   148162: contig of 3362 bp in length
            *   148163   148262: gap of unknown length
            *   148263   151505: contig of 3243 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-463I23"
BASE COUNT    45000 a  31399 c  30313 g  43563 t   1230 others
ORIGIN


  Query Match              4.6%;  Score 50;  DB 67;  Length 151505;
  Best Local Similarity   62.7%;  Pred. No. 0.39;
  Matches   94;  Conservative    0;  Mismatches   55;  Indels    1;  Gaps    1;

Qy     763 agggttggctatatcactttattgaattttgcattccttagacttttaaaatatactaat 822
           ||| |   || ||| | |   ||   |||||||| ||| | || ||||| |||   |||
Db  133351 AGGCTACTCTGTATTTCCTGCATGGGGTTTGCATTTCTTTGTCTCTTAAAGTATCTAAAT 133410

Qy     823 gtattctagtcttactctaaagacctttgatgttaaaggaatccttcatttatttcatat 882
           ||  |     || ||| ||| |||||||||  ||   ||||     ||  | ||||| |
Db  133411 GTCATGCCAGCTAACTTTAATGACCTTTGAGATTGTTGGAAAATGGCACCT-TTTCAAAA 133469

Qy     883 tccctatctcatagggccacaattatttta 912
             |  | |||| ||||||||||  | || ||
Db  133470 ACTTTCTCTCTTAGGGCCACACATTTTCTA 133499


RESULT   8
AC006417
LOCUS       AC006417   159066 bp    DNA             PRI       29-JAN-1999
DEFINITION  Homo sapiens, clone hRPK.9_A_1, complete sequence.
ACCESSION   AC006417
VERSION     AC006417.1  GI:4204181
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159066)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone hRPK.9_A_1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159066)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 159066)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 29, 1999 this sequence version replaced gi:4185159.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .159066
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.9_A_1"
                     /clone_lib="RPCI-11 human BAC library"
     repeat_region   104. .560
                     /rpt_family="MLT2FB"
     repeat_region   complement(900. .1016)
                     /rpt_family="FLAM_A"
     repeat_region   1803. .1866
                     /rpt_family="AT_rich"
     repeat_region   2257. .2523
                     /rpt_family="MIR"
     repeat_region   2678. .2928
                     /rpt_family="MER2"
     repeat_region   complement(2929. .3227)
                     /rpt_family="AluSq"
     repeat_region   3228. .3314
                     /rpt_family="MER2"
     repeat_region   3981. .4014
                     /rpt_family="AT_rich"
     repeat_region   4464. .4628
                     /rpt_family="L2"
     repeat_region   4638. .4680
                     /rpt_family="(TGAA)n"
     repeat_region   4748. .4896
                     /rpt_family="MIR"
     repeat_region   5435. .5486
                     /rpt_family="CT-rich"
     repeat_region   5505. .5664
                     /rpt_family="MIR"
     repeat_region   6546. .6567
                     /rpt_family="AT_rich"
     repeat_region   complement(6698. .6832)
                     /rpt_family="L2"
     repeat_region   complement(7558. .7855)
                     /rpt_family="AluSx"
     repeat_region   complement(8679. .8995)
                     /rpt_family="AluSx"
     repeat_region   9483. .9507
                     /rpt_family="(TAG)n"
     repeat_region   complement(9513. .9715)
                     /rpt_family="MIR"
     repeat_region   complement(11799. .12054)
                     /rpt_family="MER96B"
     repeat_region   12082. .12104
                     /rpt_family="AT_rich"
     repeat_region   complement(12138. .12327)
                     /rpt_family="MIR"
     repeat_region   12796. .13107
                     /rpt_family="MER33"
     repeat_region   13216. .13492
                     /rpt_family="AluSc"
```

```
repeat_region   13493. .13526
                /rpt_family="(CAAA)n"
repeat_region   13527. .13537
                /rpt_family="AluSc"
repeat_region   complement(13845. .14201)
                /rpt_family="MLT1A1"
repeat_region   complement(14202. .14298)
                /rpt_family="MER5A"
repeat_region   14690. .14755
                /rpt_family="L2"
repeat_region   complement(15741. .15794)
                /rpt_family="L2"
repeat_region   15854. .16077
                /rpt_family="MIR"
repeat_region   16090. .16116
                /rpt_family="AT_rich"
repeat_region   complement(16143. .16280)
                /rpt_family="L2"
repeat_region   16510. .16732
                /rpt_family="L2"
repeat_region   17116. .17252
                /rpt_family="MIR"
repeat_region   18712. .18794
                /rpt_family="C-rich"
repeat_region   complement(18795. .19087)
                /rpt_family="AluJo"
repeat_region   19173. .19539
                /rpt_family="(TA)n"
repeat_region   complement(19567. .19599)
                /rpt_family="Alu"
repeat_region   complement(19626. .21766)
                /rpt_family="L1MA6"
repeat_region   21779. .21912
                /rpt_family="(TGG)n"
repeat_region   21977. .22399
                /rpt_family="L2"
repeat_region   22444. .22810
                /rpt_family="L2"
repeat_region   complement(22850. .22908)
                /rpt_family="MIR"
repeat_region   complement(23111. .23495)
                /rpt_family="MSTD"
repeat_region   complement(23533. .23617)
                /rpt_family="AluJ/FLAM"
repeat_region   complement(23712. .24257)
                /rpt_family="L1MB8"
repeat_region   24258. .24557
                /rpt_family="AluSx"
repeat_region   complement(24558. .24770)
                /rpt_family="L1MB8"
repeat_region   complement(24919. .25161)
                /rpt_family="L1MB8"
repeat_region   25232. .25251
                /rpt_family="AT_rich"
repeat_region   25252. .25545
                /rpt_family="AluSg"
repeat_region   25546. .25569
                /rpt_family="AT_rich"
repeat_region   25820. .26006
                /rpt_family="MIR"
repeat_region   26820. .27196
                /rpt_family="L2"
repeat_region   27712. .27751
                /rpt_family="AT_rich"
repeat_region   complement(28472. .28698)
                /rpt_family="MIR"
repeat_region   complement(29696. .29896)
                /rpt_family="L2"
repeat_region   30100. .30220
                /rpt_family="L1MB8"
repeat_region   30280. .30608
                /rpt_family="MLT2E"
repeat_region   30612. .30634
                /rpt_family="(TC)n"
repeat_region   30651. .30711
                /rpt_family="AT_rich"
repeat_region   30712. .30760
                /rpt_family="MLT2E"
repeat_region   complement(31541. .31596)
                /rpt_family="L2"
repeat_region   complement(31681. .31983)
                /rpt_family="AluJo"
repeat_region   complement(33463. .33892)
                /rpt_family="MLT1K"
repeat_region   34209. .34232
                /rpt_family="(T)n"
repeat_region   35092. .35125
                /rpt_family="AT_rich"
repeat_region   complement(35156. .35518)
                /rpt_family="L1MC4"
repeat_region   complement(35519. .35818)

  Query Match             4.6%;  Score 50;  DB 86;  Length 159066;
  Best Local Similarity   62.7%;  Pred. No. 0.39;
  Matches   94;  Conservative    0;  Mismatches   55;  Indels    1;  Gaps    1;

Qy    763 agggttggctatatcactttattgaattttgcattccttagacttttaaaatatactaat 822
          ||| |   || |||  | |   ||   |||||||| ||| | || ||||| |||   |||
Db 106965 AGGCTACTCTGTATTTCCTGCATGGGGTTTGCATTTCTTTGTCTCTTAAAGTATCTAAAT 107024

Qy    823 gtattctagtcttactctaaagacctttgatgttaaaggaatccttcatttatttcatat 882
          ||  |     || ||| ||| |||||||||  ||   ||||     ||  | ||||| |
Db 107025 GTCATGCCAGCTAACTTTAATGACCTTTGAGATTGTTGGAAAATGGCACCT-TTTCAAAA 107083

Qy    883 tccctatctcatagggccacaattatttta 912
            |  | |||| ||||||||||  | || ||
Db 107084 ACTTTCTCTCTTAGGGCCACACATTTTCTA 107113


RESULT   9
AC073886/c
LOCUS       AC073886   165567 bp    DNA             HTG       20-OCT-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-101N13, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC073886
VERSION     AC073886.3  GI:10800311
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165567)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165567)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Oct 13, 2000 this sequence version replaced gi:8979965.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0101N13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 69%
            Sequencing vector: plasmid; 7%
            Chemistry: Dye-primer ET; 69% of reads
            Chemistry: Dye-terminator Big Dye; 7% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 161255 bases at least Q40
```

```
          Consensus quality: 162696 bases at least Q30
          Consensus quality: 163422 bases at least Q20
          Insert size: 176000; agarose-fp
          Insert size: 164667; sum-of-contigs
          Quality coverage: 7.47 in Q20 bases; agarose-fp
          Quality coverage: 6.47 in Q20 bases; sum-of-contigs
          ----------------------------------------------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     2128: contig of 2128 bp in length
          *     2129     2228: gap of unknown length
          *     2229    37677: contig of 35449 bp in length
          *    37678    37777: gap of unknown length
          *    37778    42500: contig of 4723 bp in length
          *    42501    42600: gap of unknown length
          *    42601    47806: contig of 5206 bp in length
          *    47807    47906: gap of unknown length
          *    47907    55928: contig of 8022 bp in length
          *    55929    56028: gap of unknown length
          *    56029    63861: contig of 7833 bp in length
          *    63862    63961: gap of unknown length
          *    63962    76511: contig of 12550 bp in length
          *    76512    76611: gap of unknown length
          *    76612   101033: contig of 24422 bp in length
          *   101034   101133: gap of unknown length
          *   101134   125861: contig of 24728 bp in length
          *   125862   125961: gap of unknown length
          *   125962   165567: contig of 39606 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165567
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-101N13"
     misc_feature    1. .2128
                     /note="assembly_name:Contig1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2229. .37677
                     /note="assembly_name:Contig10
                     clone_end:T7
                     vector_side:left"
     misc_feature    37778. .42500
                     /note="assembly_name:Contig2"
     misc_feature    42601. .47806
                     /note="assembly_name:Contig3"
     misc_feature    47907. .55928
                     /note="assembly_name:Contig4"
     misc_feature    56029. .63861
                     /note="assembly_name:Contig5"
     misc_feature    63962. .76511
                     /note="assembly_name:Contig6"
     misc_feature    76612. .101033
                     /note="assembly_name:Contig7"
     misc_feature    101134. .125861
                     /note="assembly_name:Contig8"
     misc_feature    125962. .165567
                     /note="assembly_name:Contig9"
BASE COUNT    52115 a  30967 c  30511 g  51068 t    906 others
ORIGIN


  Query Match             4.6%;  Score 49.4;  DB 75;  Length 165567;
  Best Local Similarity   58.5%;  Pred. No. 0.53;
  Matches   86;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy    745 ttttttctttcttaggtgaagggttggctatatcactttattgaattttgcattccttaga 804
```

```
          | ||   ||| |  ||  |  |||| || ||  |||| |||| |||||  ||| ||| |
Db 118689 TATTCTATTCATTTGTAGAATGTTGTCTGTAATACTTGATTGGATTTTCTATTTCTTTGG 118630

Qy    805 cttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaaggaat 864
          || |||| |||    |||   || |   |   || |||||| | ||||  |  | ||| |
Db 118629 CTGTTAAGATAATTAAATACCTTTTCACCAAGCTTTAAAGATCCTTGAACTATATGGACT 118570

Qy    865 ccttcatttatttcatattccctatct 891
           |||  | | | ||   ||   | | |
Db 118569 TCTTTCTCTCTCTCTCTTTTTTTTTTT 118543
```

```
RESULT  10
CNS06FEG/c
LOCUS       CNS06FEG      896 bp    DNA             STS       10-JAN-2001
DEFINITION  T3 end of clone AR0AA029E01 of library AR0AA from strain CBS 732 of
            Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION   AL396302
VERSION     AL396302.1  GI:12148291
KEYWORDS    STS.
SOURCE      Zygosaccharomyces rouxii.
  ORGANISM  Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE   1  (bases 1 to 896)
  AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 8.
            Zygosaccharomyces rouxii(1)
  JOURNAL   FEBS Lett. 487 (1), 52-55 (2000)
   PUBMED   11152883
REFERENCE   2  (bases 1 to 896)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 896)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .896
                     /organism="Zygosaccharomyces rouxii"
                     /strain="CBS 732"
                     /db_xref="taxon:4956"
                     /clone="AR0AA029E01"
                     /clone_lib="AR0AA"
                     /note="end : T3"
     misc_feature    <1. .>896
                     /note="part of rDNA repeats
                     contains 35S rDNA gene"
                     /evidence=not_experimental
BASE COUNT      312 a    132 c    162 g    199 t     91 others
ORIGIN
```

```
  Query Match             4.5%;  Score 48.6;  DB 53;  Length 896;
  Best Local Similarity   43.9%;  Pred. No. 0.82;
  Matches   90;  Conservative   23;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    743 tgtttttctttcttaggtgaagggttggctatatcactttattgaattttgcattcctta 802
          | :|||||||: ||  ||      |   || |||    || |:   ||||   |   |:|
Db    572 TTYTTTTCTTYATTTWGTTTTTTTTCCCCTTTATTTTATTTTWTCYTTTTYTTTGTTTWA 513

Qy    803 gacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaagga 862
            ||||     : |  |  | | || || |  : ||||||     ||| :  ||  |
Db    512 CTCTTTCTTTTWTTCTTTTTTTTTTTTTAATTCWTCTCTAATTGKTTTTTWAATTCCATCT 453

Qy    863 atccttcatttatttcatattccctatctcatagggccacaattattttaatacagagat 922
          :    | : |  :|||   :| |||:| || ||     :|: || |:| |:|:   | ||
Db    452 WAAGCTWWATACWTTCTAKWTACCTWTTTCTTACAWTTWCWTTTTTKTAAWTWTTAAAAT 393

Qy    923 gattttcaaaatattttaacaactg 947
            | ||| ||::||::|| | :| |:
Db    392 AACTTTTAAWWTAKWTTKAAWAATK 368


RESULT  11
I66494
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Schelflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             4.5%;  Score 48.4;  DB 10;  Length 7218;
  Best Local Similarity   4.8%;  Pred. No. 0.89;
  Matches   19;  Conservative  214;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    556 ctacattttagtttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatcc 615
          :: : ::::  ::::::::::::::::: : :: :  :   : :    : :::    :::
Db   1074 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1133

Qy    616 agtgcttataaacctttccttcctttgtgcacagaatgtaactagcaagcccattagcac 675
          : ::: :   :::::::::::::: : : :   : :  ::  :   ::: ::  : :
Db   1134 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1193

Qy    676 ccagataattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaagttc 735
          ::   :  :::: :: : ::  :::::: :::      ::::: : :  :      :::
Db   1194 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1253

Qy    736 cccgatgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgca 795
          :::  : : ::::::::::::   :      ::  :: : :: :::: ::   :::: :
Db   1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313

Qy    796 ttccttagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgt 855
          ::::::   :::::    : : ::  : : ::::  :::: ::::     :::::  : :
Db   1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373

Qy    856 taaaggaatccttcatttatttcatattccctatctcatagggccacaattattttaata 915
          :       :::::: ::: :::: : :::::: :::: :    :: :  :: ::::  :
Db   1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433

Qy    916 cagagatgattttcaaaatattttaacaactggtacag 953
          :  |    ||| |   |    |||||| ||| | | ||
Db   1434 YGTACCAAATTCTTCTATCTCTTTAACTACTTGCATAG 1471


RESULT  12
AC079785/c
LOCUS       AC079785   182336 bp    DNA             HTG       18-OCT-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-414P16, WORKING DRAFT
            SEQUENCE, 36 unordered pieces.
ACCESSION   AC079785
VERSION     AC079785.4  GI:10864246
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182336)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182336)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Oct 18, 2000 this sequence version replaced gi:10717322.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0414P16
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 162516 bases at least Q40
            Consensus quality: 168738 bases at least Q30
            Consensus quality: 171701 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 178836; sum-of-contigs
            Quality coverage: 3.90 in Q20 bases; agarose-fp
            Quality coverage: 3.96 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 36 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1590: contig of 1590 bp in length
            *     1591     1690: gap of unknown length
            *     1691     3539: contig of 1849 bp in length
            *     3540     3639: gap of unknown length
            *     3640     5330: contig of 1691 bp in length
            *     5331     5430: gap of unknown length
            *     5431     6649: contig of 1219 bp in length
            *     6650     6749: gap of unknown length
            *     6750    11601: contig of 4852 bp in length
            *    11602    11701: gap of unknown length
            *    11702    14154: contig of 2453 bp in length
            *    14155    14254: gap of unknown length
            *    14255    16316: contig of 2062 bp in length
            *    16317    16416: gap of unknown length
            *    16417    18009: contig of 1593 bp in length
            *    18010    18109: gap of unknown length
            *    18110    19940: contig of 1831 bp in length
            *    19941    20040: gap of unknown length
```

```
            *    20041    21644: contig of 1604 bp in length
            *    21645    21744: gap of unknown length
            *    21745    23592: contig of 1848 bp in length
            *    23593    23692: gap of unknown length
            *    23693    25815: contig of 2123 bp in length
            *    25816    25915: gap of unknown length
            *    25916    27801: contig of 1886 bp in length
            *    27802    27901: gap of unknown length
            *    27902    30880: contig of 2979 bp in length
            *    30881    30980: gap of unknown length
            *    30981    34951: contig of 3971 bp in length
            *    34952    35051: gap of unknown length
            *    35052    37192: contig of 2141 bp in length
            *    37193    37292: gap of unknown length
            *    37293    40283: contig of 2991 bp in length
            *    40284    40383: gap of unknown length
            *    40384    43838: contig of 3455 bp in length
            *    43839    43938: gap of unknown length
            *    43939    47318: contig of 3380 bp in length
            *    47319    47418: gap of unknown length
            *    47419    51525: contig of 4107 bp in length
            *    51526    51625: gap of unknown length
            *    51626    55198: contig of 3573 bp in length
            *    55199    55298: gap of unknown length
            *    55299    60177: contig of 4879 bp in length
            *    60178    60277: gap of unknown length
            *    60278    66683: contig of 6406 bp in length
            *    66684    66783: gap of unknown length
            *    66784    71458: contig of 4675 bp in length
            *    71459    71558: gap of unknown length
            *    71559    75920: contig of 4362 bp in length
            *    75921    76020: gap of unknown length
            *    76021    81481: contig of 5461 bp in length
            *    81482    81581: gap of unknown length
            *    81582    85861: contig of 4280 bp in length
            *    85862    85961: gap of unknown length
            *    85962    91757: contig of 5796 bp in length
            *    91758    91857: gap of unknown length
            *    91858    96034: contig of 4177 bp in length
            *    96035    96134: gap of unknown length
            *    96135   103395: contig of 7261 bp in length
            *   103396   103495: gap of unknown length
            *   103496   112114: contig of 8619 bp in length
            *   112115   112214: gap of unknown length
            *   112215   120240: contig of 8026 bp in length
            *   120241   120340: gap of unknown length
            *   120341   128842: contig of 8502 bp in length
            *   128843   128942: gap of unknown length
            *   128943   140927: contig of 11985 bp in length
            *   140928   141027: gap of unknown length
            *   141028   158719: contig of 17692 bp in length
            *   158720   158819: gap of unknown length
            *   158820   182336: contig of 23517 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182336
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-414P16"
     misc_feature    1. .1590
                     /note="assembly_name:Contig14"
     misc_feature    1691. .3539
                     /note="assembly_name:Contig15"
     misc_feature    3640. .5330
                     /note="assembly_name:Contig16"
     misc_feature    5431. .6649
                     /note="assembly_name:Contig17"
     misc_feature    6750. .11601
                     /note="assembly_name:Contig41
                     clone_end:T7
                     vector_side:left"
     misc_feature    11702. .14154
                     /note="assembly_name:Contig18"
     misc_feature    14255. .16316
                     /note="assembly_name:Contig19"
     misc_feature    16417. .18009
                     /note="assembly_name:Contig20"
     misc_feature    18110. .19940
                     /note="assembly_name:Contig21"
     misc_feature    20041. .21644
                     /note="assembly_name:Contig22"
     misc_feature    21745. .23592
                     /note="assembly_name:Contig23"
     misc_feature    23693. .25815
                     /note="assembly_name:Contig24"
     misc_feature    25916. .27801
                     /note="assembly_name:Contig25"
     misc_feature    27902. .30880
                     /note="assembly_name:Contig26"
     misc_feature    30981. .34951
                     /note="assembly_name:Contig27"
     misc_feature    35052. .37192
                     /note="assembly_name:Contig28"
     misc_feature    37293. .40283
                     /note="assembly_name:Contig29"
     misc_feature    40384. .43838
                     /note="assembly_name:Contig30"
     misc_feature    43939. .47318
                     /note="assembly_name:Contig31"
     misc_feature    47419. .51525
                     /note="assembly_name:Contig32"
     misc_feature    51626. .55198
                     /note="assembly_name:Contig33"
     misc_feature    55299. .60177
                     /note="assembly_name:Contig34"
     misc_feature    60278. .66683
                     /note="assembly_name:Contig35"
     misc_feature    66784. .71458
                     /note="assembly_name:Contig36"
     misc_feature    71559. .75920
                     /note="assembly_name:Contig37"
     misc_feature    76021. .81481
                     /note="assembly_name:Contig38"
     misc_feature    81582. .85861
                     /note="assembly_name:Contig39"
     misc_feature    85962. .91757
                     /note="assembly_name:Contig40"
     misc_feature    91858. .96034
                     /note="assembly_name:Contig42"
     misc_feature    96135. .103395
                     /note="assembly_name:Contig43"
     misc_feature    103496. .112114
                     /note="assembly_name:Contig44"
     misc_feature    112215. .120240
                     /note="assembly_name:Contig45"
     misc_feature    120341. .128842
                     /note="assembly_name:Contig46"
     misc_feature    128943. .140927
                     /note="assembly_name:Contig47"
     misc_feature    141028. .158719
                     /note="assembly_name:Contig48"
     misc_feature    158820. .182336

  Query Match             4.5%;  Score 48;  DB 77;  Length 182336;
  Best Local Similarity   50.0%;  Pred. No. 1.1;
  Matches  120;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy     452 atggatttcttttggttgacaaacattctggctctcagatgcaaaaagtcacactgggaa 511
           ||||||| | ||| || | ||| | | ||| || ||| |
Db   16042 ATGGATTGGGTGTGGAAGAAGGTTCATATGGAATTGTAAGTGAAAGTAACATACTTTAAC 15983

Qy     512 atgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagttttt 571
           ||  |||  | | ||| || |  ||| | |||       | ||| | | |
Db   15982 TCAAAAAATAAAATATAATATTTTTTCTTCAATTATAATTATTTTCTTCTCTTATTGTCT 15923

Qy     572 ctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgcttataaacctt 631
```

```
          |||||  |||  ||    ||| || ||||| | |  | |||| |   |    |||  ||
Db  15922 ATCTTCATTTCATTGAGAAAATTAAAGTAAAATTTAAGAGATCAAAACCAAGAAAAGATT 15863

Qy    632 tccttcctttgtgcacagaatgtaactagcaagcccattagcacccagataattctatca 691
          |  ||| |||||| ||| |||||      |   | ||   | ||    || ||| |
Db  15862 ATTATTATTTTTGCACAAAATCTAACTTCAGGACTAGTCAGTTACTAGCCCTTTTTATTA 15803


RESULT  13
CNS07E8R
LOCUS       CNS07E8R      964 bp    DNA             STS       12-JAN-2001
DEFINITION  T7 end of clone XBD0AA002B08 of library XBD0AA from strain CBS 94
            of Candida tropicalis, sequence tagged site.
ACCESSION   AL441457
VERSION     AL441457.1  GI:12224683
KEYWORDS    STS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
   PUBMED   11152891
REFERENCE   2  (bases 1 to 964)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="XBD0AA002B08"
                     /clone_lib="XBD0AA"
                     /note="end : T7"
     STS             1. .964
BASE COUNT      269 a    106 c     78 g    425 t     86 others
ORIGIN


  Query Match             4.4%;  Score 47.8;  DB 54;  Length 964;
  Best Local Similarity   37.1%;  Pred. No. 1.2;
  Matches  131;  Conservative   52;  Mismatches  164;  Indels    6;  Gaps    1;

Qy    543 atttaatttaaaactacattttagtttttctcttctcttctatgttgcaatgaatgtaaa 602
          | ||  |   | | | :| |:|  ||||| | || | || ||| |   |:| :|: ::|

Db    157 AATTTTTGATAGATTTYAATYTTTTTTTTTTTTTTTTTTTTTATTTCTAAWTTWAWWWWAC 216

Qy    603 gtatttgggatccagtgcttataaaccttttccttcctttgtgcacagaatgtaactagca 662
            :|   |:    |  || ||    |:|  || :|:|  :  :|| |:| |   :   :
Db    217 TACAYTYTAAWWWTTTTTTTTTATTWTTWTTTTTTYTWTTWWTWCACAWTTTTTTWTWWW 276

Qy    663 agcccattagcacccagataattctatcatgttagtttcccatcctggaaaatct----- 717
              ||::  :  || ::: |: | |  | |||  :| | :|     |:: | :
Db    277 TWWTAATWWTTWTWCAAWYWTTYTTTTTTTTTTTACAWTACAWTWACAWAWWTTWYAGACA 336

Qy    718 -ttgtacagtgggaagttccccgatgtgtttttctttcttaggtgaagggttggctatat 776
           |: || | |   :: :|      | |  :|:| ||| ||:|||      ||   : | |
Db    337 ATKTTAAATTTAAWWTYTTTTAWTTTTAAWTWTWTTTTTTWGGTTTTYTWTTCATWTTTT 396

Qy    777 cactttattgaattttgcattccttagacttttaaaatatactaatgtattctagtctta 836
            |:| ::   ||:|   ||  ||   |::|:  :|   || |:: | || :  |:||
Db    397 TYTTYTYWWWTTTTWTATTTTWTTTTTTCWWTYTTWACYAACAAWWTTTTTTWTTTYTTT 456

Qy    837 ctctaaagacctttgatgttaaaggaatccttcatttatttcatattccctat 889
            | |    |: |||  | | |    | |  ||  ||| |||::| |||  | |
Db    457 TTTTTTTTAYTTTTCTTTTCACCTCATTTTTTTTYTTTTTTTYWTTTTCTTTTT 509


RESULT  14
AC011061
LOCUS       AC011061   147155 bp    DNA             HTG       12-MAR-2000
DEFINITION  Homo sapiens clone RP11-12H18, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC011061
VERSION     AC011061.4  GI:7230073
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147155)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-12H18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147155)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6649411.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3268
                Center clone name: 12_H_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
```

```
            Consensus quality: 141604 bases at least Q40
            Consensus quality: 144655 bases at least Q30
            Consensus quality: 145613 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 146155; sum-of-contigs
            Quality coverage: 4.9 in Q20 bases; agarose-fp
            Quality coverage: 5.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1276: contig of 1276 bp in length
            *    1277 1376: gap of      100 bp
            *    1377     5581: contig of 4205 bp in length
            *    5582 5681: gap of      100 bp
            *    5682    10096: contig of 4415 bp in length
            *   10097 10196: gap of      100 bp
            *   10197    14496: contig of 4300 bp in length
            *   14497 14596: gap of      100 bp
            *   14597    22481: contig of 7885 bp in length
            *   22482 22581: gap of      100 bp
            *   22582    31479: contig of 8898 bp in length
            *   31480 31579: gap of      100 bp
            *   31580    44748: contig of 13169 bp in length
            *   44749 44848: gap of      100 bp
            *   44849    61899: contig of 17051 bp in length
            *   61900 61999: gap of      100 bp
            *   62000    86375: contig of 24376 bp in length
            *   86376 86475: gap of      100 bp
            *   86476   110819: contig of 24344 bp in length
            *  110820 110919: gap of      100 bp
            *  110920   147155: contig of 36236 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-12H18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1276
                     /note="assembly_fragment"
     misc_feature    1377. .5581
                     /note="assembly_fragment"
     misc_feature    5682. .10096
                     /note="assembly_fragment"
     misc_feature    10197. .14496
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    14597. .22481
                     /note="assembly_fragment"
     misc_feature    22582. .31479
                     /note="assembly_fragment"
     misc_feature    31580. .44748
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    44849. .61899
                     /note="assembly_fragment"
     misc_feature    62000. .86375
                     /note="assembly_fragment"
     misc_feature    86476. .110819
                     /note="assembly_fragment"
     misc_feature    110920. .147155
                     /note="assembly_fragment"
BASE COUNT    41235 a  32908 c  30919 g  41077 t   1016 others
ORIGIN
```

```
  Query Match             4.4%;  Score 47.8;  DB 62;  Length 147155;
  Best Local Similarity   55.0%;  Pred. No. 1.2;
  Matches   94;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     770 gctatatcacttttattgaattttgcattccttagacttttaaaatatactaatgtattct 829
           ||| || ||| | |||  |||||||    ||| |  |||| |||      ||  | ||
Db    4869 GCTCTACCACCTGATTTGATTTTGCTCATCTTTGGTTTTTCAAACTGGTAAAGTTGTTTA 4928

Qy     830 agtcttactctaaagacctttgatgttaaaggaatccttcatttatttcatattccctat 889
           |||| || ||||||  ||||||     | |||||  |  ||||  ||  |  |  || |
Db    4929 TGTCTAACACTAAAGCTCTTTGAACCAGATGGAATTCAACATTCTTTCAAATTCTACTCT 4988

Qy     890 ctcatagggccacaattattttaatacagagatgattttcaaaatatttta 940
           ||    -  |||||  |  ||  |    | | |   ||||  ||  |  | |
Db    4989 CTTTAGACTTCACAAGTCTTAGAGATAATAAAATGATTTCTGAAGAAATCA 5039


RESULT  15
AC009451
LOCUS       AC009451    182444 bp    DNA             HTG       10-FEB-2001
DEFINITION  Homo sapiens chromosome 17 clone CTD-2286H12 map 17, *** SEQUENCING
            IN PROGRESS ***, 5 unordered pieces.
ACCESSION   AC009451
VERSION     AC009451.11  GI:12740366
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182444)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2286H12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182444)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 10, 2001 this sequence version replaced gi:12658725.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L892
                Center clone name: 2286_H_12
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1    74411: contig of 74411 bp in length
            *    74412 74511: gap of      100 bp
            *    74512   133681: contig of 59170 bp in length
            *   133682 133781: gap of      100 bp
            *   133782   136258: contig of 2477 bp in length
            *   136259 136358: gap of      100 bp
            *   136359   177875: contig of 41517 bp in length
            *   177876 177975: gap of      100 bp
            *   177976   182444: contig of 4469 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182444
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2286H12"
                     /clone_lib="CITD Human BAC"
BASE COUNT    49729 a  41144 c  41855 g  49310 t    406 others
ORIGIN


  Query Match              4.4%;  Score 47.8;  DB 61;  Length 182444;
  Best Local Similarity   55.0%;  Pred. No. 1.2;
  Matches   94;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     770 gctatatcactttattgaattttgcattccttagacttttaaaatatactaatgtattct 829
           ||| || ||| | |||  |||||||    ||| |  |||| |||      ||  | ||
Db  108592 GCTCTACCACCTGATTTGATTTTGCTCATCTTTGGTTTTTCAAACTGGTAAAGTTGTTTA 108651

Qy     830 agtcttactctaaagacctttgatgttaaaggaatccttcatttatttcatattccctat 889
            |||| || ||||||  ||||||     | |||||  |  ||||  ||  |  |   || |
Db  108652 TGTCTAACACTAAAGCTCTTTGAACCAGATGGAATTCAACATTCTTTCAAATTCTACTCT 108711

Qy     890 ctcatagggccacaattattttaatacagagatgattttcaaaatatttta 940
           ||        |||||  | ||  |    | | |    ||||  || |  | |
Db  108712 CTTTAGACTTCACAAGTCTTAGAGATAATAAAATGATTTCTGAAGAAATCA 108762


Search completed: April 30, 2001, 18:54:38
Job time: 36152 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 08:53:06 ; Search time 933.78 Seconds
(without alignments)
673.945 Million cell updates/sec

Title: US-09-721-183-1
Perfect score: 1078
Sequence: 1 ggatgatacaagagccaaga.....tggtgttcctttgaatcgtg 1078

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	45.8	4.2	9789	17	T41852	cDNA encoding Plas
c	2	40.6	3.8	487	20	Z42066	Human endometrium
c	3	40.6	3.8	492	20	Z42090	Human endometrium
c	4	39.8	3.7	6888	21	A70114	Plasmodium falcipa
c	5	39.6	3.7	2340	9	N81108	Sequence of the ci
c	6	39.4	3.7	969	21	C79784	Human secreted pro
c	7	39.4	3.7	85680	21	F22299	BAC containing rep
c	8	39.4	3.7	611590	21	F22303	Arabidopsis thalia
c	9	38.6	3.6	1872	21	A47169	DNA sequence encod
	10	38.2	3.5	237326	19	V57903	Hereditary haemoch
c	11	38	3.5	571	20	X20393	Borrelia burgdorfe
	12	38	3.5	617	20	X36836	Human XLIS gene fr
c	13	38	3.5	2335	9	N81781	Sequence of the ci
c	14	38	3.5	2337	7	N60362	Sequence encoding
	15	38	3.5	9542	20	X20260	Borrelia burgdorfe
	16	37.4	3.5	1664976	19	V21209	Methanococcus jann
	17	37.2	3.5	2209	18	T87427	Clone J5. Homo sa
	18	37.2	3.5	2280	19	V52604	Human metalloprote
c	19	37	3.4	802	20	X20466	Human secreted pro
c	20	36.8	3.4	18177	10	N90490	DNA of human retin
c	21	36.8	3.4	18303	20	X04502	Human retinoblasto
c	22	36.4	3.4	5139	21	A70139	Plasmodium falcipa
c	23	36.4	3.4	116277	20	X20249	Borrelia burgdorfe
	24	36.4	3.4	910715	20	X20248	Borrelia burgdorfe
c	25	36.2	3.4	605	18	T91299	Human M97-2 secret
c	26	36.2	3.4	1134	20	X99562	Nucleic acid seque
c	27	36	3.3	4590	7	N60472	Sequence encoding
	28	36	3.3	58407	19	V21210	Methanococcus jann
c	29	36	3.3	133894	17	T13635	AcNPV genomic DNA
	30	35.8	3.3	3293	11	Q05753	Microspore-specifi
	31	35.8	3.3	3293	18	T48862	Brassica napus mic
	32	35.8	3.3	14280	18	V74368	Staphylococcus aur
c	33	35.6	3.3	267	20	X99598	Nucleic acid seque
	34	35.6	3.3	999	21	A70234	Plasmodium falcipa
	35	35.6	3.3	1828	22	C88108	Human FLEXHT-39 nu
c	36	35.6	3.3	3231	21	C61658	DNA sequence of a
c	37	35.2	3.3	269	21	Z93797	Sequence encoding
c	38	35.2	3.3	859	20	X61830	B. burgdorferi ant
c	39	35.2	3.3	951	20	X61829	B. burgdorferi ant
	40	35.2	3.3	1492	21	C61579	DNA encoding a yea
	41	35.2	3.3	18359	20	X20255	Borrelia burgdorfe
	42	35	3.2	249	21	C03082	Human secreted pro
c	43	35	3.2	2908	21	A26305	Human secreted pro
c	44	35	3.2	3872	20	Z23700	P. falciparum 1-de
c	45	35	3.2	3872	21	X82995	P. falciparum 1-dc

ALIGNMENTS

RESULT 1
T41852/c
ID T41852 standard; DNA; 9789 BP.
XX
AC T41852;
XX
DT 20-FEB-1997 (first entry)
XX
DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
XX
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite; ss.
XX
OS Plasmodium falciparum MC type.
XX

FH	Key	Location/Qualifiers
FT	CDS	326..9497
FT		/*tag= a
FT		/product= Erythrocyte membrane protein
FT	misc_feature	518..520
FT		/*tag= b
FT		/transl_except= GTA encodes Tyrosine
FT	misc_feature	656..658
FT		/*tag= c
FT		/transl_except= ATT encodes Leucine
FT	misc_feature	2909..2911
FT		/*tag= d
FT		/transl_except= AAC encodes Aspartic acid
FT	misc_feature	3461..3463
FT		/*tag= e
FT		/transl_except= GAA encodes Glutamine
FT	misc_feature	5546..5548
FT		/*tag= f
FT		/transl_except= CCT encodes Arginine

```
FT   misc_feature    6254..6256
FT                   /*tag= g
FT                   /transl_except= AAT encodes Lysine
FT   misc_feature    6257..6259
FT                   /*tag= h
FT                   /transl_except= ATA encodes Tyrosine
FT   misc_feature    6263..6265
FT                   /*tag= i
FT                   /transl_except= AAC encodes Lysine
FT   misc_feature    6269..6271
FT                   /*tag= j
FT                   /transl_except= TTC encodes Isoleucine
FT   misc_feature    6272..6274
FT                   /*tag= k
FT                   /transl_except= ATA encodes Histidine
FT   misc_feature    6275..6277
FT                   /*tag= l
FT                   /transl_except= ATT encodes Asparagine
FT   misc_feature    6278..6280
FT                   /*tag= m
FT                   /transl_except= GGA encodes Tryptophan
FT   intron          7754..8478
FT                   /*tag= n
XX
PN   WO9633736-A1.
XX
PD   31-OCT-1996.
XX
PF   26-APR-1996;   96WO-US05798.
XX
PR   27-APR-1995;   95US-0430908.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Baruch DI,  Howard RJ,  Pasloske BL;
XX
DR   WPI; 1996-497376/49.
DR   P-PSDB; W00384.
XX
PT   New Plasmodium falciparum erythrocyte membrane proteins - used to
PT   develop products for the diagnosis, treatment or prevention of
PT   malaria parasite infections
XX
PS   Disclosure; Figure 12; 149pp; English.
XX
CC   A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC   membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection.  The polypeptides can inhibit, block or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC   probes and primers to identify a Plasmodium falciparum parasite, the
CC   primers used to generate characteristic amplification patterns from
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes the
CC   PfEMP1 protein of the MC type of Plasmodium falciparum.  An
CC   alternative, truncated version of the coding sequence (a cDNA clone)
CC   is given in T41853.
XX
SQ   Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T; 0 other;

  Query Match             4.2%;  Score 45.8;  DB 17;  Length 9789;
  Best Local Similarity   46.5%;  Pred. No. 0.049;
  Matches  216;  Conservative    0;  Mismatches  247;  Indels    2;  Gaps    2;

Qy    475 cattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtgaaatta 534
Db   8362 CATTATGTATATAATATATATATATATTATATGTGTATATGTATGTGTATGAATAAAATA 8303

Qy    535 gttttggtatttaatttaaaactacattttagtttttctcttctcttctatgttgcaatg 594
Db   8302 TTTTTTTTTTTTTTTTTTTTTAATTTTTTTATTTTTTTTTTTTATTTTAT-TTTTTTTA 8244

Qy    595 aatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacagaatgt 654
Db   8243 ATTTTTTTATTTTTTTAATTTTTTTTTTTTTTTACATTTTTTTTATTTTTTTTTATTTTAT 8184

Qy    655 aactagcaagcccattagcacccagataattctatcatgttagtttcccatcctggaaaa 714
Db   8183 ATATATATATATATTTTTTTTTAACATTTTTTTAAATTTTTTTTTTTATTTTATAATATA 8124

Qy    715 tctttgtacagtgggaagttccccgatgtgtttttctttcttaggtgaagggttggctat 774
Db   8123 TATATTTCTTTTAACATTTTTTTTTATTTTATGATATAT-ATATATATATTTTTTAACAT 8065

Qy    775 atcactttattgaattttgcattccttagacttttaaaatatactaatgtattctagtct 834
Db   8064 TTTTTTTTCTTTTTTTTTTTATTTAATAAATTTTTTTTTATTTTATTTCTAATAAATTTT 8005

Qy    835 tactctaaagacctttgatgttaaaggaatccttcatttatttcatattccctatctcat 894
Db   8004 TTTTATAATCATTATTTTTATTAATATAATTTTTTTTTTATAAAATTTTTCATTTCTTTT 7945

Qy    895 agggccacaattattttaatacagagatgattttcaaaatatttt 939
Db   7944 TTTTTTATGAATTTTTTTTTTTTTATTTATTTTAATAAATTTTTT 7900


RESULT   2
Z42066/c
ID   Z42066 standard; cDNA; 487 BP.
XX
AC   Z42066;
XX
DT   31-JAN-2000  (first entry)
XX
DE   Human endometrium tumour cDNA derived EST 86.
XX
KW   Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:
KW   treatment; uterine; gene therapy; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   DE19817948-A1.
XX
PD   21-OCT-1999.
XX
PF   17-APR-1998;   98DE-1017948.
XX
PR   17-APR-1998;   98DE-1017948.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591957/51.
DR   P-PSDB; Y60198, Y60199, Y60200, Y60201.
XX
PT   New nucleic acid sequences expressed in uterine cancer tissues, and
PT   derived polypeptides, for treatment of uterine and endometrial cancer
PT   and identification of therapeutic agents  -
XX
PS   Claim 3; Page 231; 444pp; German.
XX
CC   This invention describes novel human nucleic acid (cDNA) sequences (A),
CC   that are highly expressed in uterine tumour tissue and which have
CC   anticancer and cytostatic activity. (A) are used (i) for recombinant
CC   expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC   are used (i) to identify agents suitable for treatment of uterine or
CC   endometrial cancer; (ii) directly for treating these forms of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
```

```
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. Z41981-Z42121 represent EST fragments
CC   derived from a human endometrium tumour cDNA library which encode the
CC   protein sequences represented in Y59941-Y60328.
XX
SQ   Sequence 487 BP; 133 A; 106 C; 106 G; 142 T; 0 other;


  Query Match             3.8%;  Score 40.6;  DB 20;  Length 487;
  Best Local Similarity   73.2%;  Pred. No. 0.38;
  Matches   52;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy     931 aaatattttaacaactggtacaggacagatgccagccactcagaagggatgcctgctgta 990
           |||   |||||||||||||||| || | |||  ||| ||| ||||||  ||||  | ||  ||
Db     351 AAAATGTTTAACAACTGGTATAGCATAGACTCCAACCAATCAGAATGGATTTCAGCCATA 292

Qy     991 aacaagcagta 1001
           |||   |||||
Db     291 AACTGCCAGTA 281


RESULT   3
Z42090/c
ID   Z42090 standard; cDNA; 492 BP.
XX
AC   Z42090;
XX
DT   31-JAN-2000  (first entry)
XX
DE   Human endometrium tumour cDNA derived EST 110.
XX
KW   Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:
KW   treatment; uterine; gene therapy; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   DE19817948-A1.
XX
PD   21-OCT-1999.
XX
PF   17-APR-1998;   98DE-1017948.
XX
PR   17-APR-1998;   98DE-1017948.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591957/51.
DR   P-PSDB; Y60275, Y60276, Y60277.
XX
PT   New nucleic acid sequences expressed in uterine cancer tissues, and
PT   derived polypeptides, for treatment of uterine and endometrial cancer
PT   and identification of therapeutic agents -
XX
PS   Claim 3; Page 248; 444pp; German.
XX
CC   This invention describes novel human nucleic acid (cDNA) sequences (A),
CC   that are highly expressed in uterine tumour tissue and which have
CC   anticancer and cytostatic activity. (A) are used (i) for recombinant
CC   expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC   are used (i) to identify agents suitable for treatment of uterine or
CC   endometrial cancer; (ii) directly for treating these forms of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. Z41981-Z42121 represent EST fragments
CC   derived from a human endometrium tumour cDNA library which encode the
CC   protein sequences represented in Y59941-Y60328.
XX
SQ   Sequence 492 BP; 134 A; 106 C; 107 G; 145 T; 0 other;


  Query Match             3.8%;  Score 40.6;  DB 20;  Length 492;
  Best Local Similarity   73.2%;  Pred. No. 0.38;
  Matches   52;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy     931 aaatattttaacaactggtacaggacagatgccagccactcagaagggatgcctgctgta 990
           |||   |||||||||||||||| || | |||  ||| ||| ||||||  ||||  | ||  ||
Db     351 AAAATGTTTAACAACTGGTATAGCATAGACTCCAACCAATCAGAATGGATTTCAGCCATA 292

Qy     991 aacaagcagta 1001
           |||   |||||
Db     291 AACTGCCAGTA 281


RESULT   4
A70114/c
ID   A70114 standard; DNA; 6888 BP.
XX
AC   A70114;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:247.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 467-469; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
```

```
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 6888 BP; 3027 A; 533 C; 1088 G; 2240 T; 0 other;


  Query Match             3.7%;  Score 39.8;  DB 21;  Length 6888;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  104;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy     680 ataattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaagttccccg 739
           ||||||   | |   ||  |   | | |   |||   |       |   |||| | |. |
Db    4735 ATAATTTGGTTAACATATATAGTCTTTCATTAAATAATCCTATAAAATGGAAATGCTTCA 4676

Qy     740 atgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgcattcc 799
           |||| || |    | |    | ||  |||   ||| |    |||     ||||    |
Db    4675 ATGTTTTCTCACATTTCTCATTAAATGTTTTATATTTAGGATTAACACCTTTTAATATTA 4616

Qy     800 ttagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaa 859
            || |||||||  |||||   || || | | | | | || |  |||| ||| ||  ||
Db    4615 ATATACTTTTATTATATAAAGATATAATATTGCCATTCTGTGCAGACATTTTATAATATT 4556

Qy     860 ggaatccttcatttatttcatattccctatc 890
             ||  ||   |||  |  | |    ||||
Db    4555 TATATGATTTTCTTAAATTCTCTATTTTATC 4525


RESULT   5
N81108/c
ID   N81108 standard; DNA; 2340 BP.
XX
AC   N81108;
XX
DT   20-NOV-1990  (first entry)
XX
DE   Sequence of the circumsporozoite (CS) gene from Plasmodium falciparum.
XX
KW   Vaccine; antigen; immunogen; probe; hybridisation;
KW   immunoassay; diagnosis; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             78..1316
FT                   /*tag=  a
FT   CDS             396..440
FT                   /*tag=  b
FT                   /note="Region 1"
FT   repeat_region   495..698
FT                   /*tag=  c
FT                   /rpt_unit=AAT GCA AAC CCA
FT   repeat_region   447..494
FT                   /*tag=  d
FT                   /rpt_unit=AAT GCA AAN CCA AAT GTA GAT CCC
FT   repeat_region   711..938
FT                   /*tag=  e
FT                   /rpt_unit=AAT GCA AAT CCA
XX
PN   EP278940-A.
XX
PD   17-AUG-1988.
XX
PF   25-JAN-1988;   88EP-0870008.
XX
PR   30-JAN-1987;   87US-0009325.
XX
PA   (SMIK ) SMITH KLINE-RIT.
XX
PI   Cabezon T,  De Wilde M,  Harford N;
XX
DR   WPI; 1988-229751/33.
DR   P-PSDB; P80508.
XX
PT   DNA encoding hepatitis B virus antigens and hybrids contg. them -
PT   used for expression in yeast to obtain vaccines and bivalent
PT   vaccines
XX
PS   Example; Fig 3Aa-3Af; 101pp; English.
XX
CC   Sequence of the CS gene (n81108) is from lambda-mPfl. A recombinant DNA
CC   molecule is claimed, comprising functional DNA coding sequence fused, in
CC   phase, to a portion of the Pre-S2 region of a hepatitis B virus (HBV)
CC   Pre-S2-S protein coding sequence. The functional DNA coding sequence
CC   comprises the Pre-S2 coding sequence, Pre-S1 coding sequence or entire
CC   Pre-S1-Pre-S2 coding sequence, the CS protein coding sequence of
CC   Plasmodium, or an HIV coding sequence such as
CC   an HIV envelope gene sequence, e.g. HIV C7 protein coding region, HIV
CC   Peptide 121 coding region, or HIV Dreesman peptide coding region.
XX
SQ   Sequence 2340 BP; 1027 A; 323 C; 308 G; 681 T; 1 other;


  Query Match             3.7%;  Score 39.6;  DB 9;  Length 2340;
  Best Local Similarity   45.1%;  Pred. No. 1.1;
  Matches  147;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy     613 tccagtgcttataaaccttccttcctttgtgcacagaatgtaactagcaagcccattag 672
           ||||    | || ||      ||||  || |    |  ||   | ||  |     ||||
Db    2149 TCCAAAAATAATTAAGTCCAACTTCTCTTATCTGTATTATAATATTATTATTATTATTAT 2090

Qy     673 cacccagataattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaag 732
           |     || ||  |||  |  |   ||    |  | |   ||  ||   |  |  |
Db    2089 TATTACAATCATCATATTCTTCTTTCTTTATTTTATCGTTTATTATTAATCCATATGTCT 2030

Qy     733 ttccccgatgtgtttttctttcttaggtgaagggttggctatatcactttattgaatttt 792
           || |  ||  | |||||  ||| |   |     | |         |   ||   |  |||||
Db    2029 TTTCTTGAATTATTTTTGCTTCATGAATATTTGTTGTAAATCCTATATTGGATTCATTTT 1970

Qy     793 gcattccttagacttttaaaatatactaatgtattctagtcttactctaaagacctttga 852
           |||  |||     ||  ||  ||  | ||  ||| |||| ||       |  ||   |
Db    1969 TTATTTCTTCTGTATTATAATCATTATTATCAATTTTAGTGTTCACACTACCACAAATAT 1910

Qy     853 tgttaaaggaatccttcatttatttcatattccctatctcatagggccacaattatttta 912
           | ||| |   |   ||  | ||||  |||   | |||          |||||  |||  ||
Db    1909 TATTACAATTAATATTGTTGTATTATATAAAGCGTATACGTCTTATGCACAAATATAGTA 1850

Qy     913 atacagagatgattttcaaaatattt 938
           |  ||  | ||  ||  |||| |
Db    1849 CATTTGTTATTTTATTTAATATATCT 1824


RESULT   6
C79784/c
ID   C79784 standard; cDNA; 969 BP.
XX
AC   C79784;
XX
DT   12-FEB-2001  (first entry)
XX
DE   Human secreted protein gene 45 SEQ ID NO:55.
XX
KW   Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; infection;
KW   wound healing; food additive; preservative; skin aging; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058494-A1.
XX
```

```
PD   05-OCT-2000.
XX
PF   23-MAR-2000; 2000WO-US07578.
XX
PR   26-MAR-1999;   99US-0126507.
PR   07-JAN-2000; 2000US-0174872.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594644/56.
DR   P-PSDB; B44739.
XX
PT   New nucleic acid molecules encoding 50 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 336-337; 371pp; English.
XX
CC   The polynucleotide sequences given in C79740 to C79789 encode the human
CC   secreted proteins given in B44695 to B44744. B44745 to B44760 represent
CC   human secreted polypeptide sequences and proteins homologous to them,
CC   which are given in the exemplification of the present invention. Human
CC   secreted proteins have activities based on the tissues and cells the
CC   genes are expressed in. Examples of activities include: antiarthritic;
CC   immunosuppressive; antirheumatic; antiproliferative; cytostatic;
CC   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
CC   antibacterial; virucide; fungicide; ophthalmological; and vulnerary. The
CC   polynucleotides and polypeptides can be used to prevent, treat or
CC   ameliorate a medical condition in e.g. humans, mice, rabbits, goats,
CC   horses, cats, dogs, chickens or sheep. They are also used in diagnosing
CC   a pathological condition or susceptibility to a pathological condition.
CC   Disorders which are diagnosed or treated include autoimmune diseases,
CC   hyperproliferative disorders, cardiovascular disorders, cerebrovascular
CC   disorders, angiogenesis, nervous system disorders, infections caused by
CC   bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC   also be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. C79731 to C79739
CC   and B44694 represent sequences used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 969 BP; 276 A; 204 C; 216 G; 271 T; 2 other;

  Query Match             3.7%;  Score 39.4;  DB 21;  Length 969;
  Best Local Similarity   51.4%;  Pred. No. 0.95;
  Matches   91;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     534 agttttggtatttaatttaaaactacattttagtttttctcttctcttctatgttgcaat 593
           | | | | || ||| |||| |  ||  |  |||||  |  |||||||| | || ||||
Db     917 ATTCTAGATAATTACTTTAGATGTATCTCGTAGTTCATTGCTTCTCTTTTCTGCTGCATC 858

Qy     594 gaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacagaatg 653
             ||  ||| || |   | |||||  ||     |  || ||   ||  | ||  || |   |
Db     857 TAACTTAATGTTTAATGTATCCATTGAGATTTTACTTTCAGTTATTGTGCTCATATTCTT 798

Qy     654 taactagcaagcccattagcacccagataattctatcatgttagtttcccatcctgg 710
             |||    |   || |   | |   |  | ||| |    |  | |  || |  ||
Db     797 ACACTTCTTACTGTATCACTTCTCTTCTGCTCCTAACTCCATCTTATATCACCAGGG 741


RESULT   7
F22299/c
ID   F22299 standard; DNA; 85680 BP.
XX
AC   F22299;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #22.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 763-777; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 85680 BP; 27189 A; 16414 C; 15246 G; 26711 T; 120 other;

  Query Match             3.7%;  Score 39.4;  DB 21;  Length 85680;
  Best Local Similarity   55.5%;  Pred. No. 4.2;
  Matches   76;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     510 aaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagttt 569
           | | |||    || || ||| | || | | |   | |  | ||  |||||||| ||||||
Db   72500 AGAAGAAGGAGAACTGATGAGACTAATGTAGAGCTCTGTTGTACTACTACATTATAGTTT 72441

Qy     570 ttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgcttataaacc 629
           ||  | ||| |   |  |||  || |    |   |  || |  || | |   ||||   |
Db   72440 TTACCATCTTTAAAAGTTTGGTATTATAAGAGGTTCGTTTGATTCAATTTTCTATATTTC 72381

Qy     630 tttccttcctttgtgca 646
           ||| || || | || |
Db   72380 ATTCTTTGCTATATGAA 72364


RESULT   8
F22303/c
ID   F22303 standard; DNA; 611590 BP.
XX
AC   F22303;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Arabidopsis thaliana chromosome 2 centromere.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
```

PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0153584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited michrosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells -
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;

Query Match 3.7%; Score 39.4; DB 21; Length 611590;
Best Local Similarity 55.5%; Pred. No. 8.2;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 510 aaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagttt 569
Db 598730 AGAAGAAGGAGAACTGATGAGACTAATGTAGAGCTCTGTTGTACTACTACATTATAGTTT 598671

Qy 570 ttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgcttataaacc 629
Db 598670 TTACCATCTTTAAAAGTTTGGTATTATAAGAGGTTCGTTTGATTCAATTTTCTATATTTC 598611

Qy 630 tttccttcctttgtgca 646
Db 598610 ATTCTTTGCTATATGAA 598594

RESULT 9
A47169/c
ID A47169 standard; DNA; 1872 BP.
XX
AC A47169;
XX
DT 03-OCT-2000 (first entry)
XX
DE DNA sequence encoding Streptococcus protective antigen Spa18.
XX
KW spa18 gene; Streptococcus protective antigen; surface antigen;
KW group A Streptococci; Streptococcus infection; ss.
XX
OS Streptococcus sp.
XX
PN WO200037648-A1.
XX
PD 29-JUN-2000.
XX
PF 22-DEC-1999; 99WO-US30807.
XX
PR 22-DEC-1998; 98US-0113794.
PR 31-DEC-1998; 98US-0114730.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Dale J;
XX
DR WPI; 2000-442675/38.
XX
PT New Spa polypeptides and polynucleotides isolated from a Streptococcus
PT species, useful for eliciting opsonic antibodies in an animal that are
PT protective against Streptococci infections -
XX
PS Claim 26; Fig 5A-C; 85pp; English.
XX
CC The present sequence encodes a Streptococcus protective antigen.
CC Spa is a surface antigen distinct from M protein which evokes
CC opsonic antibodies that are protective against multiple serotypes
CC of streptococci. Spa is an antigen of group A Streptococci. The
CC spa18 polypeptides, nucleic acid molecules and antibodies are
CC useful for protecting an animal against a Streptococcus infection.
XX
SQ Sequence 1872 BP; 731 A; 287 C; 384 G; 469 T; 1 other;

Query Match 3.6%; Score 38.6; DB 21; Length 1872;
Best Local Similarity 54.6%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 796 ttccttagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgt 855
Db 689 TTTTTTTCAGATTTTAAATCAACTTCTGCATTCTTACGTTGCTCAAGTTCTGCAGTCTGT 630

Qy 856 taaaggaatccttcatttatttcatattccctatctcatagggccacaattattttaata 915
Db 629 TTAAGTAATTCTTCCTTTAGATGTGACTCACTATCTAAACGGTATTTAATTGCTTCAATC 570

Qy 916 cagagatgattttcaaaatat 936
Db 569 ACTAAACTTCCTAAAGAATAT 549

RESULT 10
V57903
ID V57903 standard; DNA; 237326 BP.
XX
AC V57903;
XX
DT 21-DEC-1998 (first entry)
XX
DE Hereditary haemochromatosis subregion from an HH affected individual.
XX
KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
XX
OS Homo sapiens.
XX
PN WO9814466-A1.
XX
PD 09-APR-1998.
XX
PF 30-SEP-1997; 97WO-US17658.
XX
PR 07-MAY-1997; 97US-0852495.
PR 01-OCT-1996; 96US-0724394.
XX
PA (PROG-) PROGENTIOR INC.
XX
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ;
PI Tsuchihashi Z, Wolff RK;
XX
DR WPI; 1998-240014/21.

```
XX
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
XX
PS   Claim 1; Fig 9; 209pp; English.
XX
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an hereditary
CC   haemochromatosis (HH) affected individual. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
XX
SQ   Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T; 0 other;


  Query Match             3.5%;  Score 38.2;  DB 19;  Length 237326;
  Best Local Similarity   54.0%;  Pred. No. 12;
  Matches  101;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
           |||||  |  |       |  ||  | |  |||| |  | | || |  | | | | |
Db  126329 acaaatgtaatcctatgtacctgacatatttcactccagaacattaggtttgaatagatt 126388

Qy     530 aattagttttggtatttaatttaaaactacattttagtttttctcttctcttctatgttg 589
           ||  || ||| | | ||       | | |||||  || || | |  | ||| |||||
Db  126389 catctgtgttgctgtgtataacttaattcatttttattgttatgtaatattccatgtt- 126447

Qy     590 caatgaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacag 649
             |||| || ||    || ||  || | || | ||  |  ||| ||||| || | ||
Db  126448 --atgagtgcaacaatttaggtgtctactgttgatgcatatttgcttccctttttcagct 126505

Qy     650 aatgtaa 656
           ||| |||
Db  126506 aatataa 126512


RESULT  11
X20393/c
ID   X20393 standard; DNA; 571 BP.
XX
AC   X20393;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #146.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 1109; 1128pp; English.
XX
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 571 BP; 198 A; 76 C; 54 G; 241 T; 2 other;


  Query Match             3.5%;  Score 38;  DB 20;  Length 571;
  Best Local Similarity   46.9%;  Pred. No. 1.8;
  Matches  119;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy     691 atgttagtttcccatcctggaaaatctttgtacagtgggaagttccccgatgtgtttttc 750
           || | | |||   ||  || |||  |||  | | | ||  ||    | | | ||
Db     520 ATCTAATTTTAGAATTTTGTAAATAATTTAGAGAATAGGTTTTTATAAGTTCTTTTAAAA 461

Qy     751 tttcttaggtgaagggttggctatatcactttattgaattttgcattccttagacttttta 810
           |    |     ||  | |   ||| |   | | || |            | | | | |||
Db     460 TAAGATTTTATAAAAGCTTTATATTTTTGTATTTTTATAGACCGCAGTGTAATAATATTA 401

Qy     811 aaatatactaatgtattctagtcttactctaaagacctttgatgttaaaggaatccttca 870
           |   ||  |||| ||   | |   ||  |||||  |   ||        ||  |||  ||||
Db     400 ATTGATTTTAATTTAAGGTTGAACTAAACTAAATATAGTTTTGTAGGAAATAATTTTTCA 341

Qy     871 tttatttcatattccctatctcatagggccacaattattttaatacagagatgattttca 930
           ||  || |  ||   |||   || |    |  ||||      |||  || |  |||  |
Db     340 TTATTTCCTACTTGAATATTGGATCGTAAAAATATTAGGGCTTTACTAAGTTCTTTTAAA 281

Qy     931 aaatattttaacaa 944
           | | | |||| |||
Db     280 AGAGAATTTAGCAA 267


RESULT  12
X36836
ID   X36836 standard; DNA; 617 BP.
XX
AC   X36836;
XX
DT   14-JUL-1999  (first entry)
XX
DE   Human XLIS gene fragment including intron cos 4.
XX
KW   XLIS gene; human; detection; diagnosis; prenatal diagnosis; therapy;
KW   lissencephaly; LIS; agyria-pachygyria; subcortical laminar heterotopia;
KW   SCLH; cortical dysgenesis; cryptogenic epilepsy; neurological disorder;
KW   neurodegenerative disease; Alzheimer's disease; X-linked disorder;
KW   genetic counselling; ss.
XX
```

```
OS   Homo sapiens.
XX
PN   EP918091-A1.
XX
PD   26-MAY-1999.
XX
PF   21-NOV-1997;   97EP-0402811.
XX
PR   21-NOV-1997;   97EP-0402811.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Chelly J, Des Portes V, Kahn A, Pinard J;
XX
DR   WPI; 1999-290318/25.
XX
PT   New gene and its gene product expressed in the brain, useful for
PT   diagnosing and treating disorders such as lissencephaly and
PT   subcortical laminar heterotopia
XX
PS   Claim 1; Page 24-25; 71pp; English.
XX
CC   This sequence is a fragment of the human XLIS gene of the invention.
CC   The XLIS fragments may be used to detect abnormalities in the expression
CC   of the XLIS gene transcripts or to compare their sequence with that of
CC   the XLIS transcripts from patients for in vitro especially prenatal
CC   diagnosis of lissencephaly (LIS) (or agyria-pachygyria), subcortical
CC   laminar heterotopia (SCLH), cortical dysgenesis, cryptogenic epilepsies
CC   or neurodegenerative diseases such as Alzheimer's disease. These
CC   disorders mainly affect females as the XLIS gene is X-linked. The XLIS
CC   fragments may also be used to administer to patients to prevent or treat
CC   the above disorders and may be used as a tool in genetic counselling.
CC   Oligonucleotides which bind to the fragments may be used to amplify the
CC   XLIS gene from a sample for comparison to normal samples in the in vitro
CC   diagnosis regime. This may also be performed by amplifying XLIS cDNA from
CC   the mRNA in the sample. Antibodies to XLIS may be used to detect XLIS in
CC   a biological sample or can be administered to patients to prevent or
CC   treat the above disorders. They may also be used to purify XLIS from a
CC   biological sample. XLIS may also be administered to patients to prevent
CC   or treat the above neurological disorders. In addition XLIS may be used
CC   as a marker of neuronal cells at an early stage of development; its
CC   discovery increases understanding of both the neuronal movement which
CC   leads to development of the cortical region of the brain and of the
CC   pathogenesis of the group of neuronal disorders mentioned above.
XX
SQ   Sequence 617 BP; 149 A; 176 C; 67 G; 218 T; 7 other;

  Query Match             3.5%;  Score 38;  DB 20;  Length 617;
  Best Local Similarity   52.5%;  Pred. No. 1.9;
  Matches   83;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     472 aaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtgaaa 531
           ||  |||  ||  | ||||  |||||   | |||     ||||| |||  |      ||
Db     198 aacaatttaggtcttaagattaaaaaaaaaatactttagtatgaagtgtctgaacaagaa 257

Qy     532 ttagttttggtatttaatttaaaactacattttagtttttctcttctcttctatgttgca 591
           |    ||||  |  |  || |    | |||||||| | || | ||| ||||   |  ||
Db     258 tccactttgaaagctgtttaattgttccattttaatattccacttttctttcttcatgat 317

Qy     592 atgaatgtaaagtatttgggatccagtgcttataaaacc 629
             ||   ||||    |    | | || |  |||||| |
Db     318 tggatattaaaaagctcattagcaagggaatataaaac 355


RESULT  13
N81781/c
ID   N81781 standard; DNA; 2335 BP.
XX
AC   N81781;
XX
DT   18-SEP-1990  (first entry)
XX
DE   Sequence of the circumsporozoite (CS) gene of Plasmodium falciparum in
DE   lambda mPfl.
XX
KW   Circumsporozoite gene; Plasmodium falciparum; lambda mPfl; vaccine;
KW   yeast glyceraldehyde-3P-dehydrogenase gene (TDH3) promoter;
KW   yeast ornithine carbamoyl transferase gene (ARG3).
XX
OS   Plasmodium falciparum .
XX
FH   Key             Location/Qualifiers
FT   CDS             78..1316
FT                   /*tag=  a
FT   misc_feature    396..440
FT                   /*tag=  b
FT                   /note="region I"
FT   misc_feature    1107..1145
FT                   /*tag=  c
FT                   /note="region II"
XX
PN   WO8805817-A.
XX
PD   11-SEP-1988.
XX
PF   25-JAN-1988;   88WO-BE00002.
XX
PR   30-JAN-1987;   87US-0008791.
XX
PA   (SMIK ) SMITH KLINE RIT SA.
XX
PI   De Wilde M,  Gathoye AM;
XX
DR   WPI; 1988-235171/33.
DR   P-PSDB; P80835.
XX
PT   Expression of P. falciparum circumsporozoite protein by yeast -
PT   using recombinant DNA vector having coding sequence linked to
PT   expression control sequence
XX
PS   Example 1; Fig 2a; 44pp; English.
XX
CC   Plasmid WR201 was obtained from the Walter Reed Army Institute of
CC   Research, and results from insertion of a 2.3 kb EcoRI fragment from
CC   lambda mPfl encoding the complete CS protein gene P. falciparum into
CC   vector pUC8.  A recombinant DNA vector is claimed, which comprises a DNA
CC   sequence contg. the coding sequence of the CS P. falciparum operatively
CC   linked to an expression control sequence.  Prefd. expression control
CC   sequences include the yeast glyceraldehyde-3P-dehydrogenase gene (TDH3)
CC   promoter and the yeast ornithine carbamoyl transferase gene (ARG3)
CC   transcription termination region.  A suitable coding sequence comprises
CC   the 1215bp StuI-RsaI fragment of WR201 contg. the P. falciparum CS
CC   protein coding sequence, minus its first 50bp.  Also claimed is a
CC   transformed host cell, a method of culturing the cell to produce CS, the
CC   protein, and a vaccine.
XX
SQ   Sequence 2335 BP; 1024 A; 327 C; 301 G; 683 T; 0 other;

  Query Match             3.5%;  Score 38;  DB 9;  Length 2335;
  Best Local Similarity   45.0%;  Pred. No. 2.9;
  Matches  143;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy     621 ttataaacctttccttcctttgtgcacagaatgtaactagcaagcccattagcacccaga 680
           | || ||      ||||  || |    |  ||   | ||  |     ||||  |     |
Db    2137 TAATTAAGTCCAACTTCTCTTATCTGTATTATAATATTATTATTATTATTATTACAA 2078

Qy     681 taattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaagttccccga 740
           | ||  |||  |  |   ||    |  | |   ||  ||   |  |  |   || |  ||
Db    2077 TCATCATATTCTTCTTTCTTTATTTTATCGTTTATTATTAATCCATATGTCTTTTCTTGA 2018

Qy     741 tgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgcattcct 800
           | |||||  ||| |   |    | |        |   ||   |  |||||  ||| ||
Db    2017 ATTATTTTTGCTTCATGAATATTTGTTGTAAATCCTATATTGGATTCATTTTTTATTTCT 1958
```

```
Qy     801 tagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaag 860
           |     ||  ||  ||  | ||  ||| |||| ||       |  ||   |  | ||| |
Db    1957 TCTGTATTATAATCATTATTATCAATTTTAGTGTTCACACTACCACAAATATTATTACAA 1898

Qy     861 gaatccttcatttatttcatattccctatctcatagggccacaattattttaatacagag 920
             |   ||  | ||||  |||   | |||          |||||.|||  ||     |
Db    1897 TTAATATTGTTGTATTATATAAAGCGTATACGTCTTATGCACAAATATAGTACATTTGTT 1838

Qy     921 atgattttcaaaatattt 938
           ||  | ||  || |||| |
Db    1837 ATTTTATTTAATATATCT 1820


RESULT 14
N60362/c
ID   N60362 standard; DNA; 2337 BP.
XX
AC   N60362;
XX
DT   13-JUN-1991 (first entry)
XX
DE   Sequence encoding CS protein of malaria parasite.
XX
KW   Sporozoite; vaccination; ds.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             78..1316
FT                   /*tag= a
FT   repeat_unit     447..458
FT                   /*tag= b
FT                   /number= 37 repeats
XX
PN   EP166410-A.
XX
PD   02-JAN-1986.
XX
PF   24-JUN-1985;    85EP-0107794.
XX
PR   26-JUN-1984;    84US-0624564.
XX
PA   (USDC ) US SEC OF COMMERCE.
PA   (USGO ) US GOVERNMENT.
PA   (USSA ) US SEC OF THE ARMY.
XX
PI   McCutchan TF, Dame JB, Williams JL, Schneider I;
XX
DR   WPI; 1986-008635/02.
DR   P-PSDB; P60416.
XX
PT   New immunologically active pure synthetic peptide(s) - used for
PT   protection against infection by malaria parasite.
XX
PS   Disclosure; Fig 2; 49pp; English.
XX
CC   The Plasmodium CS gene was used to isolate peptides capable of
CC   inducing an immune response to the parasite. Peptide antigens may
CC   be synthesised in pure form and used to generate an immune
CC   response in vaccination against malaria. The featured repeat
CC   units are claimed and must be present in copies of 2-1000.
XX
SQ   Sequence 2337 BP; 1024 A; 326 C; 303 G; 684 T; 0 other;


  Query Match             3.5%;  Score 38;  DB 7;  Length 2337;
  Best Local Similarity   45.0%;  Pred. No. 2.9;
  Matches  143;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy     621 ttataaacctttccttcctttgtgcacagaatgtaactagcaagcccattagcacccaga 680
           | || ||      ||||  || |    |  ||   | ||  |     ||||  |     |
Db    2139 TAATTAAGTCCAACTTCTCTTATCTGTATTATAATATTATTATTATTATTATTATTACAA 2080

Qy     681 taattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaagttccccga 740
           | ||  |||  |  |   ||    | | |   ||  ||  |  | |  || | ||
Db    2079 TCATCATATTCTTCTTTCTTTATTTTATCGTTTATTATTAATCCATATGTCTTTTCTTGA 2020

Qy     741 tgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgcattcct 800
             | |||||  ||| |   |    | |         |   ||   |  |||||  ||| ||
Db    2019 ATTATTTTTGCTTCATGAATATTTGTTGTAAATCCTATATTGGATTCATTTTTTATTTCT 1960

Qy     801 tagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaag 860
           |     ||  ||  ||  | ||  ||| |||| ||       |  ||   |  | ||| |
Db    1959 TCTGTATTATAATCATTATTATCAATTTTAGTGTTCACACTACCACAAATATTATTACAA 1900

Qy     861 gaatccttcatttatttcatattccctatctcatagggccacaattattttaatacagag 920
             |   ||  | ||||  |||   | |||          |||||.|||  ||     |
Db    1899 TTAATATTGTTGTATTATATAAAGCGTATACGTCTTATGCACAAATATAGTACATTTGTT 1840

Qy     921 atgattttcaaaatattt 938
           ||  | ||  || |||| |
Db    1839 ATTTTATTTAATATATCT 1822


RESULT 15
X20260
ID   X20260 standard; DNA; 9542 BP.
XX
AC   X20260;
XX
DT   04-MAY-1999 (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #13.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;    98WO-US12764.
XX
PR   03-SEP-1997;    97US-0057483.
PR   20-JUN-1997;    97US-0050359.
PR   22-JUL-1997;    97US-0053344.
PR   22-JUL-1997;    97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
PS   Claim 1; Page 920-925; 1128pp; English.
XX
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 9542 BP; 3812 A; 1160 C; 1113 G; 3457 T; 0 other;
```

Query Match 3.5%; Score 38; DB 20; Length 9542;
Best Local Similarity 46.9%; Pred. No. 4.6;
Matches 119; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy     691 atgttagtttcccatcctggaaaatctttgtacagtgggaagttccccgatgtgtttttc 750
           || | | ||| || || ||| ||| | | | || || | | | ||
Db    8963 atctaattttagaattttgtaaataatttagagaataggtttttataagttcttttaaaa 9022

Qy     751 tttcttaggtgaagggttggctatatcactttattgaattttgcattccttagactttta 810
           |   |    ||  | |  ||| |  | | || |           | | | | |||
Db    9023 taagattttataaaagctttatatttttgtatttttatagaccgcagtgtaataatatta 9082

Qy     811 aaatatactaatgtattctagtcttactctaaagacctttgatgttaaaggaatccttca 870
           |  ||  |||| ||  | |  ||  ||||| |  ||      ||  |||  |||| 
Db    9083 attgattttaatttaaggttgaactaaactaaatatagttttgtaggaaataatttttca 9142

Qy     871 tttatttcatattccctatctcatagggccacaattattttaatacagagatgattttca 930
           ||  || |  ||  |||  || |  |  |||| |||  || |  |||  |
Db    9143 ttatttcctacttgaatattggatcgtaaaaatattagggctttactaagttcttttaaa 9202

Qy     931 aaatattttaacaa 944
           | | | |||| |||
Db    9203 agagaatttagcaa 9216
```

Search completed: April 30, 2001, 19:04:18
Job time: 36672 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 08:50:56 ; Search time 330.07 Seconds
(without alignments)
570.248 Million cell updates/sec

Title: US-09-721-183-1
Perfect score: 1078
Sequence: 1 ggatgatacaagagccaaga.....tggtgttcctttgaatcgtg 1078

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	48.4	4.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	39.2	3.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	3	38.2	3.5	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	4	38.2	3.5	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	5	38.2	3.5	246240	2	US-08-724-394A-22	Sequence 22, Appl
	6	37.2	3.5	2209	1	US-08-514-014-1	Sequence 1, Appli
	7	37.2	3.5	2209	2	US-08-833-823-1	Sequence 1, Appli
	8	37.2	3.5	2280	3	US-08-813-150-1	Sequence 1, Appli
	9	35.8	3.3	3293	1	US-08-030-096-1	Sequence 1, Appli
c	10	35.6	3.3	3231	4	US-08-632-806A-7	Sequence 7, Appli
	11	35.2	3.3	1492	4	US-08-350-468-7	Sequence 7, Appli
c	12	35	3.2	7295	2	US-08-487-826B-15	Sequence 15, Appl
c	13	34	3.2	1302	1	US-08-425-299A-1	Sequence 1, Appli
c	14	34	3.2	1452	1	US-08-187-785-2	Sequence 2, Appli
	15	34	3.2	5852	1	US-07-867-106-2	Sequence 2, Appli
c	16	33.8	3.1	10607	1	US-08-078-090-3	Sequence 3, Appli
c	17	33.2	3.1	152331	4	US-09-128-155-16	Sequence 16, Appl
c	18	33	3.1	4140	3	US-08-894-731-2	Sequence 2, Appli
c	19	32.8	3.0	4612	2	US-08-447-031A-8	Sequence 8, Appli
	20	32.8	3.0	87350	3	US-08-781-891-79	Sequence 79, Appl
	21	32.6	3.0	3891	1	US-08-480-604A-27	Sequence 27, Appl
	22	32.6	3.0	3891	2	US-08-405-496A-27	Sequence 27, Appl
	23	32.4	3.0	518	1	US-08-485-284A-2	Sequence 2, Appli
c	24	32.4	3.0	4084	2	US-08-568-459A-1	Sequence 1, Appli
c	25	32.4	3.0	4084	2	US-08-487-826B-1	Sequence 1, Appli
c	26	32.4	3.0	4084	6	5198347-5	Patent No. 5198347
	27	32.4	3.0	5775	1	US-08-306-691B-15	Sequence 15, Appl
	28	32.4	3.0	5775	5	PCT-US93-06251-29	Sequence 29, Appl
c	29	32.2	3.0	573	2	US-08-286-819A-20	Sequence 20, Appl
c	30	32.2	3.0	573	3	US-08-980-357-20	Sequence 20, Appl
	31	32.2	3.0	760	2	US-08-975-316-25	Sequence 25, Appl
	32	32.2	3.0	2178	3	US-08-781-891-72	Sequence 72, Appl
	33	32.2	3.0	5208	3	US-08-781-891-70	Sequence 70, Appl
c	34	32.2	3.0	10851	2	US-08-286-819A-16	Sequence 16, Appl
c	35	32.2	3.0	10851	3	US-08-980-357-16	Sequence 16, Appl
c	36	32	3.0	1991	2	US-08-415-593-40	Sequence 40, Appl
c	37	32	3.0	2128	2	US-08-415-593-39	Sequence 39, Appl
	38	31.8	2.9	971	5	PCT-US91-06234A-5	Sequence 5, Appli
c	39	31.8	2.9	176373	4	US-09-128-155-17	Sequence 17, Appl
c	40	31.6	2.9	1278	4	US-09-050-159-123	Sequence 123, App
	41	31.6	2.9	2309	3	US-09-078-317-3	Sequence 3, Appli
c	42	31.6	2.9	8224	6	5180808-1	Patent No. 5180808
c	43	31.6	2.9	37895	1	US-08-375-709-1	Sequence 1, Appli
c	44	31.6	2.9	37895	1	US-08-752-929-1	Sequence 1, Appli
c	45	31.6	2.9	37895	4	US-09-090-793-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             4.5%;  Score 48.4;  DB 1;  Length 7218;
  Best Local Similarity   4.8%;  Pred. No. 0.00071;
  Matches   19;  Conservative  214;  Mismatches  165;  Indels    0;  Gaps    0;

Qy     556 ctacattttagtttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatcc 615
           :: : ::::  :::::::::::::::::: : :: :  :   : :    : :::    :::
Db    1074 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1133

Qy     616 agtgcttataaaccttttccttcctttgtgcacagaatgtaactagcaagcccattagcac 675
             : ::: :   :::::::::::::: : : :    : :  :: :   ::: ::  : :
Db    1134 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1193

Qy     676 ccagataattctatcatgttagtttcccatcctggaaaatctttgtacagtgggaagttc 735
           ::   :  :::: :: : ::  :::::: ::::      ::::: : :  :      :::
Db    1194 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1253

Qy     736 cccgatgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgca 795
           :::  : : ::::::::::::: :      ::  :: : :: :::: ::   :::: :
Db    1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313

Qy     796 ttccttagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgt 855
           ::::::   :::::    : : ::  : : ::::  :::: ::::     :::::  : :
Db    1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373

Qy     856 taaaggaatccttcatttatttcatattccctatctcatagggccacaattattttaata 915
           :       :::::: ::: :::: : :::::: :::: :    :: :  :: ::::  :
Db    1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433

Qy     916 cagagatgattttcaaaatattttaacaactggtacag 953
           : |   ||| |   |   |||||| ||| | | ||
Db    1434 YGTACCAAATTCTTCTATCTCTTTAACTACTTGCATAG 1471


RESULT   2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             3.6%;  Score 39.2;  DB 1;  Length 7218;
  Best Local Similarity   1.9%;  Pred. No. 0.23;
  Matches    5;  Conservative  159;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       4 tgatacaagagccaagaagggacatttgagttgtgtcgcttagataggaaagggatccag 63
           || |||:::::  :::::::::: :   :::  : :  :   ::: ::::::::::   ::
Db    1440 TGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381

Qy      64 ggaaaatcaacagtaagtgaggatgagcagcgtctcttggttttcattgaggatagagta 123
           ::::::  :: :: ::: ::::: ::: :: :      ::     :  ::::: :::: :
Db    1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321

Qy     124 agagattgagtttagattgcaacagaaggaattagtttagataccaggaagaacttccta 183
           :::::  :::   :::  : :: :::::::::  ::   ::: :  ::::::::      :
Db    1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261

Qy     184 gcctgaagatttgtcatagtgtctgctttctagatatctgggaaagatttgataatagtt 243
           :   :::::   :  : :: :   :      ::: :   :::::::::   :: :: ::
Db    1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201

Qy     244 gtttgtgaatagaaaggaggatatga 269
           :   : ::: :::::::::::: : ::
Db    1200 RRRRRRRRRRRRRRRRRRRRRRRRR 1175


RESULT   3
US-08-724-394A-20/c
; Sequence 20, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
```

```
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-20


  Query Match             3.5%;  Score 38.2;  DB 2;  Length 246240;
  Best Local Similarity   54.0%;  Pred. No. 1.9;
  Matches  101;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy    470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
          |||||  |  |        |  ||  | |  |||| |  | | ||  |  | | | | |
Db 118471 ACAAATGTAATCCTATGTACCTGACATATTTCACTCCAGAACATTAGGTTTGAATAGATT 118412

Qy    530 aattagttttggtatttaatttaaaaactacattttagtttttctcttctcttctatgttg 589
           ||  || ||| | | ||       | | ||||||  || || | |  | ||| |||||
Db 118411 CATCTGTGTTGCTGTGTATAACTTTAATTCATTTTTATTGTTATGTAATATTCCATGTT- 118351

Qy    590 caatgaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacag 649
            |||| || ||    || ||  || | || | ||  |  ||| ||||| || | ||
Db 118352 --ATGAGTGCAACAATTTAGGTGTCTACTGTTGATGCATATTTGCTTCCCTTTTTCAGCT 118295

Qy    650 aatgtaa 656
          ||| |||
Db 118294 AATATAA 118288


RESULT   4
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
```

```
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-21


  Query Match             3.5%;  Score 38.2;  DB 2;  Length 246240;
  Best Local Similarity   54.0%;  Pred. No. 1.9;
  Matches  101;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy    470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
          |||||  |  |        |  ||  | |  |||| |  | | ||  |  | | | | |
Db 118471 ACAAATGTAATCCTATGTACCTGACATATTTCACTCCAGAACATTAGGTTTGAATAGATT 118412

Qy    530 aattagttttggtatttaatttaaaaactacattttagtttttctcttctcttctatgttg 589
           ||  || ||| | | ||       | | ||||||  || || | |  | ||| |||||
Db 118411 CATCTGTGTTGCTGTGTATAACTTTAATTCATTTTTATTGTTATGTAATATTCCATGTT- 118351

Qy    590 caatgaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacag 649
            |||| || ||    || ||  || | || | ||  |  ||| ||||| || | ||
Db 118352 --ATGAGTGCAACAATTTAGGTGTCTACTGTTGATGCATATTTGCTTCCCTTTTTCAGCT 118295

Qy    650 aatgtaa 656
          ||| |||
Db 118294 AATATAA 118288


RESULT   5
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
```

```
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-22


  Query Match             3.5%;  Score 38.2;  DB 2;  Length 246240;
  Best Local Similarity   54.0%;  Pred. No. 1.9;
  Matches  101;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy    470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
          |||||  |  |       |  ||  | |  |||| |  | | || |  | | | | |
Db 118471 ACAAATGTAATCCTATGTACCTGACATATTTCACTCCAGAACATTAGGTTTGAATAGATT 118412

Qy    530 aattagttttggtatttaatttaaaaactacattttagtttttctcttctcttctatgttg 589
          ||  || ||| | | ||       | | ||||||  || || | |  | ||| |||||
Db 118411 CATCTGTGTTGCTGTGTATAACTTTAATTCATTTTTATTGTTATGTAATATTCCATGTT- 118351

Qy    590 caatgaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacag 649
            |||| || ||    || ||  || | || | ||  |  ||| ||||| || | ||
Db 118352 --ATGAGTGCAACAATTTAGGTGTCTACTGTTGATGCATATTTGCTTCCCTTTTTCAGCT 118295

Qy    650 aatgtaa 656
          ||| |||
Db 118294 AATATAA 118288


RESULT   6
US-08-514-014-1
; Sequence 1, Application US/08514014
; Patent No. 5707829
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/514,014
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2209 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  38..1447
US-08-514-014-1


  Query Match             3.5%;  Score 37.2;  DB 1;  Length 2209;
  Best Local Similarity   57.9%;  Pred. No. 0.5;
  Matches   66;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    508 ggaaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagt 567
          || ||  |  ||||| | | || ||| | | |    ||     |||   | | ||||| |
Db   1596 GGTAAACAGATGTAATTAGAGACATTGGCTCTTTGTTTAGGCCTAATCTTTCTTTTTACT 1655

Qy    568 ttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgct 621
          |||| | |||| || | | ||    | ||  | | | ||||| ||   ||| ||
Db   1656 TTTTTTTTTCTTTTTTCTTTTTTTTTAAAGATCATGAATTTGTGACTTAGTTCT 1709


RESULT   7
US-08-833-823-1
; Sequence 1, Application US/08833823
; Patent No. 5969093
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn-Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,823
;      FILING DATE:  10-APR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/514,014
;      FILING DATE:  11-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2209 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  38..1447
```

US-08-833-823-1

Query Match 3.5%; Score 37.2; DB 2; Length 2209;
Best Local Similarity 57.9%; Pred. No. 0.5;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy    508 ggaaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagt 567
          || ||  |  ||||| | | || ||| | | |    ||     |||   | | ||||| |
Db   1596 GGTAAACAGATGTAATTAGAGACATTGGCTCTTTGTTTAGGCCTAATCTTTCTTTTTACT 1655

Qy    568 ttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgct 621
          |||| | |||| || | | ||    | ||  | | | ||||| ||   ||| ||
Db   1656 TTTTTTTTTCTTTTTTCTTTTTTTTTAAAGATCATGAATTTGTGACTTAGTTCT 1709
```

RESULT 8
US-08-813-150-1
; Sequence 1, Application US/08813150
; Patent No. 6069229
; GENERAL INFORMATION:
; APPLICANT: Mueller, Christopher
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Liu, Yong-Jun
; APPLICANT: Dowling, Lynette M.
; APPLICANT: Huffine, Constance M.
; APPLICANT: Gorman, Daniel M.
; TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,150
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0693
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-852-9196
; TELEFAX: 650-496-1200
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1470
US-08-813-150-1

Query Match 3.5%; Score 37.2; DB 3; Length 2280;
Best Local Similarity 57.9%; Pred. No. 0.51;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy    508 ggaaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattttagt 567
          || ||  |  ||||| | | || ||| | | |    ||     |||   | | ||||| |
Db   1618 GGTAAACAGATGTAATTAGAGACATTGGCTCTTTGTTTAGGCCTAATCTTTCTTTTTACT 1677

Qy    568 ttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgct 621
          |||| | |||| || | | ||    | ||  | | | ||||| ||   ||| ||
Db   1678 TTTTTTTTTCTTTTTTCTTTTTTTTTAAAGATCATGAATTTGTGACTTAGTTCT 1731
```

RESULT 9
US-08-030-096-1
; Sequence 1, Application US/08030096
; Patent No. 5426041
; GENERAL INFORMATION:
; APPLICANT: Fabijanski, Steven F.
; APPLICANT: Arnison, Paul G.
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID
; TITLE OF INVENTION: SEED PRODUCTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030,096
; FILING DATE: 22-MAR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/556,917
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA91/00255
; FILING DATE: 22-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/164/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3293 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(3..125, 690..1796, 1913..2209)
US-08-030-096-1

Query Match 3.3%; Score 35.8; DB 1; Length 3293;
Best Local Similarity 52.3%; Pred. No. 1.4;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy    461 ttttggttgacaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgt 520
          ||||  || |    | | |  || |    |   ||| | || ||  | ||   || ||
Db    197 TTTTTTTTCATTTTTAATATAACTGTTCTTTTACAAAGGCCAAACGAGAAATGCAAATGG 256

Qy    521 aagtggtgaaattagttttggtatttaatttaaaactacattttagtttttctcttctct 580
          ||       | || ||||     ||| ||||| |  ||   | |  ||||| | |||
Db    257 AAAGTTCTCATTTCGTTTGATATTTTCATTTACAGTTAACATATGATTTTTTTTTTCAGA 316

Qy    581 tctatgttgcaatgaatgtaaagtatttggg 611
          ||| ||| |   ||  || |||   ||| ||
Db    317 TCTTTGTAGGTTTGTTTGAAAAAAGTTTTGG 347
```

```
RESULT 10
US-08-632-806A-7/c
; Sequence 7, Application US/08632806A
; Patent No. 6127179
;  GENERAL INFORMATION:
;    APPLICANT:  DellaPenna, Dean
;    APPLICANT:  Watson, Colin F.
;    TITLE OF INVENTION:  Gene Promoter for Tomato Fruit
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53703
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,806A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  920214.90123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3231 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum
US-08-632-806A-7

  Query Match             3.3%;  Score 35.6;  DB 4;  Length 3231;
  Best Local Similarity   52.7%;  Pred. No. 1.6;
  Matches   77;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     468 tgacaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggt 527
           | |  ||  |||  | | | | |||  ||||||  | |   | | ||  ||  ||||  |  |
Db    1842 TAATTAATGTTCATGTTTTAATATGTTAAAAGTTATAAGTGAACATAAATTGTATTTTCT 1783

Qy     528 gaaattagttttggtatttaatttaaaactacattttagtttttctcttctcttctatgt 587
            ||| || ||   | ||  || | || |    |  |||   | |   |  | || |||||
Db    1782 TAAAGTATTTGACGAATAAAAATGAATAAGGAAAATTATAGTATAATTGTTTTTTTATGT 1723

Qy     588 tgcaatgaatgtaaagtatttgggat 613
           | |||       |||  || |   ||
Db    1722 TACAAGATTATGAAAAGATGTACAAT 1697


RESULT 11
US-08-350-468-7
; Sequence 7, Application US/08350468
; Patent No. 6127158
;  GENERAL INFORMATION:
;    APPLICANT:  Jentsch, Stefan
;    APPLICANT:  Kirschner, Marc W.
;    APPLICANT:  King, Randall W.
;    APPLICANT:  Yew, P. Renee
;    TITLE OF INVENTION:  Ubiquitin Conjugating Enzymes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KEVIN M. FARRELL, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII(Text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/350,468
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kevin M. Farrell
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1492 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  524..562
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  563..672
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  673..1104
US-08-350-468-7

  Query Match             3.3%;  Score 35.2;  DB 4;  Length 1492;
  Best Local Similarity   50.6%;  Pred. No. 1.5;
  Matches   85;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     532 ttagttttggtatttaatttaaaactacattttagtttttctcttctcttctatgttgca 591
           ||  || |  |  | | |  | |   | || ||  || | |  || |||   |  ||
Db    1227 TTGTTTCTTATGCTAAGTAAATAGAAAGATATTTTTTGTACCATTTTCTATAAGTATGGC 1286

Qy     592 atgaatgtaaagtatttgggatccagtgcttataaacctttccttcctttgtgcacagaa 651
           |   || || | |-|     ||| | || |||  |   ||| | | | || || |  |||
Db    1287 AACTATATACACTTTAATAAATCTATTGGTTAGTAGAATTTTCATTCATTTTGTAGTGAA 1346

Qy     652 tgtaactagcaagcccattagcacccagataattctatcatgttagtt 699
           || |||||| | |  | ||  |    ||       |  | ||||||
Db    1347 TGAAACTAGCCAACGTAGTAAAGCAATCATGGCATCTTTCTTTTAGTT 1394


RESULT 12
US-08-487-826B-15/c
; Sequence 15, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
```

```
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-15


  Query Match             3.2%;  Score 35;  DB 2;  Length 7295;
  Best Local Similarity   52.4%;  Pred. No. 3.3;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     792 tgcattccttagacttttaaaatatactaatgtattctagtcttactctaaagacctttg 851
           ||  ||   | || |||||||||| || |    ||| ||   | | |||  |     | |
Db    4527 TGTTTTATCTGGATTTTTAAAATCTATTTTATTATCCTCTCCATTCTCATATCCATTATC 4468

Qy     852 atgttaaaggaatccttcatttatttcatattccctatctcatagggccacaattatttt 911
           || || |   |||||   |||  |  | ||||   | ||    ||| | |   || | |
Db    4467 ATTTTTACATGATCCTAACTTTTGTAAAAATTCTGCAGCTGTAGGGGACGCTCCTAGTGT 4408

Qy     912 aatacagagatgattttcaaaatattt 938
             || | |  | |||  ||  ||  ||
Db    4407 TCTAGAAAATTCATTAGCATTATTATT 4381


RESULT  13
US-08-425-299A-1/c
; Sequence 1, Application US/08425299A
; Patent No. 5726025
;  GENERAL INFORMATION:
;    APPLICANT:  Kirschner, Marc W.
;    APPLICANT:  King, Randall W.
;    APPLICANT:  Peters, Jean-Michael
;    TITLE OF INVENTION:  Assay and Reagents for Detecting Inhibitors
;    TITLE OF INVENTION:  of Ubiquitin-Dependent Degradation of
;    TITLE OF INVENTION:  Cell Cycle Regulatory Proteins
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII(text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,299A
;      FILING DATE:  20-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  HMI-014
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1302 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1299
US-08-425-299A-1


  Query Match             3.2%;  Score 34;  DB 1;  Length 1302;
  Best Local Similarity   51.3%;  Pred. No. 3;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     410 atggaccaaaatcaccttcagccatgtggtttcttcatcatcatggatttcttttggttg 469
           ||  ||||||||   ||   || |     ||||||| | | || |       || ||| |
Db     369 ATCAACCAAAATAGGCTCAGGCGAAAGTTTTTCTTCTTTAACAGGCTCAGGTTCTGGCTC 310

Qy     470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
           |    |   ||    | | |   || | | |  ||||| |   | | |  |   |
Db     309 AGGTTCTGGCTCTGGCACTGGCTCAGACACTGGCACTGGCACCAGCATAGGTACCTTTTC 250

Qy     530 aattagttttggtatttaatttaaaaactacattt 563
           ||   |||||||| ||  ||   ||  || |||
Db     249 AAGAGGTTTTGGTAGTTTTTTATCAATGACTTTT 216


RESULT  14
US-08-187-785-2/c
; Sequence 2, Application US/08187785
; Patent No. 5756476
;  GENERAL INFORMATION:
;    APPLICANT:  Epstein, Stephen
;    APPLICANT:  Unger, Ellis
;    APPLICANT:  Speir, Edith
;    TITLE OF INVENTION:  Inhibition of No. 5756476-Transformed Cell
;    TITLE OF INVENTION:  Proliferation Using Anti-Sense Oligonucleotides
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson, and Bear
;      STREET:  620 Newport Center Dr.  Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/187,785
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/07/821,415
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Altman, Daniel E.
;       REGISTRATION NUMBER:  34,115
;       REFERENCE/DOCKET NUMBER:  NIH001.001A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  714-760-0404
;       TELEFAX:  714-760-9502
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1452 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Human
;     IMMEDIATE SOURCE:
;       CLONE:  Cyclin B1 cDNA clone
US-08-187-785-2


  Query Match             3.2%;  Score 34;  DB 1;  Length 1452;
  Best Local Similarity   51.3%;  Pred. No. 3.2;
  Matches   79;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     410 atggaccaaaatcaccttcagccatgtggtttcttcatcatcatggatttcttttggttg 469
           ||  ||||||||   ||   || |     |||||||  | | || |       || ||| |
Db     458 ATCAACCAAAATAGGCTCAGGCGAAAGTTTTCTTCTTTAACAGGCTCAGGTTCTGGCTC 399

Qy     470 acaaacattctggctctcagatgcaaaaagtcacactgggaaatgaactgtaagtggtga 529
           |    |   ||  | | |  || | | |  ||||| |   | |  |  | |  | |
Db     398 AGGTTCTGGCTCTGGCACTGGCTCAGACACTGGCACTGGCACCAGCATAGGTACCTTTTC 339

Qy     530 aattagttttggtatttaatttaaaactacattt 563
           ||   ||||||||| ||  ||   ||  || |||
Db     338 AAGAGGTTTTGGTAGTTTTTTATCAATGACTTTT 305


RESULT  15
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             3.2%;  Score 34;  DB 1;  Length 5852;
  Best Local Similarity   50.5%;  Pred. No. 5.7;
  Matches  108;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy     741 tgtgtttttctttcttaggtgaagggttggctatatcactttattgaattttgcattcct 800
           | | ||||| ||| ||   |      ||    || | | ||| ||   ||||  ||  |
Db    5598 TATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTTAAATTTTTTTTTTTTTTTTTT 5657

Qy     801 tagacttttaaaatatactaatgtattctagtcttactctaaagacctttgatgttaaag 860
           ||    |||||||| || | ||  ||| ||    ||    | |||  | | |    |||
Db    5658 TATTTATTTAAAATTTAATTATTAATTTTATAAATAAAAAATAGAAATATAAGTAAAAAA 5717

Qy     861 gaatccttcatttatttcatattccctatctcatagggc-cacaattattttaatacaga 919
           |  |  ||  ||    ||||     |   || |   |  | |||||  | |||  |
Db    5718 ACAAACAACAAATAACATATATAAAAAAATACAAATAACAAATAATTAAATAAATTAAAT 5777

Qy     920 gatgattttcaaaatattttaacaactggtacag 953
           |  |||   ||  ||| |||| || |   | ||
Db    5778 AACCATTAAAAATGTATATTAATAAATTTAAAAG 5811


Search completed:  May  1, 2001, 00:31:10
Job time:  56414 sec
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 08:48:51 ; Search time 12176.5 Seconds
(without alignments)
12.940 Million cell updates/sec

Title: US-09-721-183-1
Perfect score: 1078
Sequence: 1 ggatgatacaagagccaaga.....tggtgttcctttgaatcgtg 1078

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	779.4	72.3	804	105	AL046013	AL046013 DKFZp434P
c	2	468	43.4	491	22	AI570494	AI570494 tm77c05.x
c	3	227.6	21.1	376	170	BF901189	BF901189 PM4-MT023
c	4	226.6	21.0	383	170	BF901203	BF901203 PM4-MT023
	5	117.6	10.9	528	142	BE980729	BE980729 UI-M-BG2-
	6	108	10.0	154	19	AI340919	AI340919 tb52f04.x
	7	107	9.9	115	162	BE043150	BE043150 hk48f01.y
c	8	107	9.9	144	15	AI054382	AI054382 qi64d10.x
	9	107	9.9	153	19	AI349657	AI349657 tb58b02.x
	10	107	9.9	155	18	AI289554	AI289554 qw42c04.x
c	11	107	9.9	173	102	AI792441	AI792441 qi64d10.y
	12	106.4	9.9	163	19	AI341041	AI341041 tb55h06.x
	13	106	9.8	141	22	AI611456	AI611456 tt61d04.x
	14	106	9.8	147	18	AI309341	AI309341 tb51a04.x
	15	106	9.8	150	18	AI313310	AI313310 ta92d09.x
c	16	105	9.7	143	15	AI053506	AI053506 qi68c06.x
c	17	105	9.7	174	102	AI792491	AI792491 qi68c06.y
	18	104.4	9.7	149	18	AI308430	AI308430 tb43g07.x

19	104.4	9.7	157	18	AI275251	AI275251 qw50a02.x
20	104.4	9.7	175	21	AI492706	AI492706 th51d04.x
21	103.4	9.6	154	18	AI271041	AI271041 qw51g01.x
22	103	9.6	148	18	AI305846	AI305846 qw48a06.x
23	103	9.6	163	24	AI733965	AI733965 qx23d05.x
24	103	9.6	272	110	AW071349	AW071349 xa65a08.x
25	102	9.5	152	18	AI305954	AI305954 qw48e05.x
26	102	9.5	160	18	AI308281	AI308281 tb42c11.x
27	101.4	9.4	138	19	AI349920	AI349920 tb63c08.x
28	101.4	9.4	145	19	AI340760	AI340760 tb53g03.x
29	101.4	9.4	147	18	AI312359	AI312359 ta91a01.x
30	101.4	9.4	148	18	AI270815	AI270815 qw51d02.x
31	101.4	9.4	151	18	AI275254	AI275254 qw50a05.x
32	101.4	9.4	151	18	AI308410	AI308410 tb43e10.x
33	101.4	9.4	153	18	AI308597	AI308597 tb45f11.x
34	101.4	9.4	153	21	AI492798	AI492798 th52g08.x
35	101.4	9.4	157	18	AI306757	AI306757 qw44b07.x
36	101.4	9.4	157	18	AI309794	AI309794 tb47a01.x
37	101.4	9.4	157	162	BE043384	BE043384 hk37e09.y
38	101.4	9.4	159	18	AI305589	AI305589 qw47b05.x
39	101	9.4	153	18	AI307043	AI307043 qw45e07.x
40	99.8	9.3	166	20	AI432284	AI432284 th48h04.x
41	96.4	8.9	112	143	BF055771	BF055771 7i73f06.y
42	96	8.9	119	162	BE041921	BE041921 hj18b08.y
43	95.4	8.8	143	18	AI306045	AI306045 qw79e11.x
44	95.4	8.8	143	19	AI334694	AI334694 qw85d06.x
45	94.4	8.8	110	143	BF055751	BF055751 7i73d06.y

ALIGNMENTS

```
RESULT   1
AL046013
LOCUS       AL046013      804 bp    mRNA            EST       29-FEB-2000
DEFINITION  DKFZp434P226_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434P226 5', mRNA sequence.
ACCESSION   AL046013
VERSION     AL046013.1  GI:5434108
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens (
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     EST (Koehrer, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Koehrer K
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            s1 sequence also available.
            This clone (DKFZp434P226) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434P226"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      215 a    143 c    154 g    292 t
ORIGIN
```

```
  Query Match             72.3%;  Score 779.4;  DB 105;  Length 804;
  Best Local Similarity   99.6%;  Pred. No. 3.4e-181;
  Matches  802;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy      97 ctcttggttttcattgaggatagagtaagagattgagtttagattgcaacagaaggaatt 156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCTTGGTTTTCATTGAGGATAGAGTAAGAGATTGAGTTTAGATTGCAACAGAAGGAATT 60

Qy     157 agtttagataccaggaagaacttcctagcctgaagatttgtcatagtgtctgctttctag 216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTTTAGATACCAGGAAGAACTTCCTAGCCTGAAGATTTGTCATAGTGTCTGCTTTCTAG 120

Qy     217 atatctgggaaagatttgataatagttgtttgtgaatagaaaggaggatatgatgttttt 276
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATATCTGGGAAAGATTTGATAATAGTTGTTTGTGAATAGAAAGGAGGATATGATGTTTTT 180

Qy     277 attggccattttgcgggactcttcgacttcttgctgctgtctcttgaggatacattccaa 336
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATTGGCCATTTTGCGGGACTCTTCGACTTCTTGCTGCTGTCTCTTGAGGATACATTCCAA 240

Qy     337 ttccatcctggcgagatccaagtgcttacgtactgtctccttagctgccttagagtaaac 396
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTCCATCCTGGCGAGATCCAAGTGCTTACGTACTGTCTCCTTAGCTGCCTTAGAGTAAAC 300

Qy     397 gatcatcagttcaatggaccaaaatcaccttcagccatgtggtttcttcatcatcatgga 456
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GATCATCAGTTCAATGGACCAAAATCACCTTCAGCCATGTGGTTTCTTCATCATCATGGA 360

Qy     457 tttcttttggttgacaaacattctggctctcagatgcaaaaagtcacactgggaaatgaa 516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTTCTTTTGGTTGACAAACATTCTGGCTCTCAGATGCAAAAAGTCACACTGGGAAATGAA 420

Qy     517 ctgtaagtggtgaaattagttttggtatttaatttaaaactacattt-tagtttttctct 575
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     421 CTGTAAGTGGTGAAATTAGTTTTGGTATTTAATTTAAAACTACATTTATAGTTTTTCTCT 480

Qy     576 tctcttctatgttgcaatgaatgtaaagtatttgggatccagtgcttataaacctttcct 635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCTCTTCTATGTTGCAATGAATGTAAAGTATTTGGGATCCAGTGCTTATAAACCTTTCCT 540

Qy     636 tcctttgtgcacagaatgtaactagcaagcccattagcacccagataattctatcatgtt 695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TCCTTTGTGCACAGAATGTAACTAGCAAGCCCATTAGCACCCAGATAATTCTATCATGTT 600

Qy     696 agtttcccatcctggaaaatctttgtacagtgggaagttccccgatgtgtttttctttct 755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGTTTCCCATCCTGGAAAATCTTTGTACAGTGGGAAGTTCCCCGATGTGTTTTTCTTTCT 660

Qy     756 taggtgaagggttggctatatcactttattgaattttgcattccttagacttttaaaata 815
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TAGGTGAAGGGTTGGCTATATCACTTTATTGAATTTTGCATTCCTTAGACTTTTAAAATA 720

Qy     816 tactaatgtattctagtcttactctaaagacctttgatgttaaaggaatccttcatttat 875
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     721 TACTAATGTATTCTAGTCTTACTCTAAAGACCTTTGATGTTAAAGGAATCCTTCA-TTAT 779

Qy     876 ttcatattccctatctcatagggcc 900
           ||||||||||||||||||||||| |
Db     780 TTCATATTCCCTATCTCATAGGGGC 804
```

```
RESULT   2
AI570494/c
LOCUS       AI570494      491 bp    mRNA            EST       14-APR-1999
DEFINITION  tm77c05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2164136 3',
            mRNA sequence.
ACCESSION   AI570494
VERSION     AI570494.1  GI:4533868
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1145   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2164136"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      174 a     87 c     90 g    139 t      1 others
ORIGIN
```

```
  Query Match           43.4%;  Score 468;  DB 22;  Length 491;
  Best Local Similarity  99.4%;  Pred. No. 1.1e-104;
  Matches  490;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy     586 gttgcaatgaatgtaaagtatttgggatccagtgcttataaaccttccttcctttgtgc 645
           |||||||||||||| |||||||||||| |||||||||||||| ||||||||||||||
Db     491 GTTGCAATGAATGTANAGTATTTGGGAT-CAGTGCTTATAAACC-TTCCTTCCTTTGTGC 434

Qy     646 acagaatgtaactagcaagcccattagcacccagataattctatcatgttagtttcccat 705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     433 ACAGAATGTAACTAGCAAGCCCATTAGCACCCAGATAATTCTATCATGTTAGTTTCCCAT 374

Qy     706 cctggaaaatctttgtacagtgggaagttccccgatgtgtttttctttcttaggtgaagg 765
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 CCTGGAAAATCTTTGTACAGTGGGAAGTTCCCCGATGTGTTTTTCTTTCTTAGGTGAAGG 314

Qy     766 gttggctatatcactttattgaattttgcattccttagacttttaaaatatactaatgta 825
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 GTTGGCTATATCACTTTATTGAATTTTGCATTCCTTAGACTTTTAAAATATACTAATGTA 254

Qy     826 ttctagtcttactctaaagacctttgatgttaaaggaatccttcatttatttcatattcc 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 TTCTAGTCTTACTCTAAAGACCTTTGATGTTAAAGGAATCCTTCATTTATTTCATATTCC 194

Qy     886 ctatctcatagggccacaattatttttaatacagagatgattttcaaaatattttaacaac 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 CTATCTCATAGGGCCACAATTATTTTTAATACAGAGATGATTTTCAAAATATTTTAACAAC 134

Qy     946 tggtacaggacagatgccagccactcagaagggatgcctgctgtaaacaagcagtatgta 1005
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 TGGTACAGGACAGATGCCAGCCACTCAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTA 74

Qy    1006 tggttgtaccaatgcctattggctgaacattatgctactttcagatattaaaatggtgtt 1065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 TGGTTGTACCAATGCCTATTGGCTGAACATTATGCTACTTTCAGATATTAAAATGGTGTT 14

Qy    1066 cctttgaatcgtg 1078
           |||||||||||||
Db      13 CCTTTGAATCGTG 1
```

```
RESULT   3
BF901189/c
LOCUS       BF901189      376 bp    mRNA            EST       18-JAN-2001
DEFINITION  PM4-MT0234-081200-001-b01 MT0234 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF901189
VERSION     BF901189.1  GI:12292648
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-MT0234-
            081200-001-b01&t3=2000-12-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 8
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0234"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       94 a    103 c     72 g    107 t
ORIGIN
```

```
  Query Match           21.1%;  Score 227.6;  DB 170;  Length 376;
  Best Local Similarity  98.3%;  Pred. No. 1.2e-45;
  Matches  230;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggatgatacaagagccaagaagggacatttgagttgtgtcgcttagataggaaagggatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 GGATGATACAAGAGCCAAGAAGGGACATTTGAGTTGTGTCGCTTAGATAGGAAAGGGATC 177
```

```
Qy      61 cagggaaaatcaacagtaagtgaggatgagcagcgtctcttggttttcattgaggataga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176 CAGGGAAAATCAACAGTAAGTGAGGATGAGCAGCGTCTCTTGGTTTTCATTGAGGATAGA 117

Qy     121 gtaagagattgagtttagattgcaacagaaggaattagtttagataccaggaagaacttc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 GTAAGAGATTGAGTTTAGATTGCAACAGAAGGAATTAGTTTAGATACCAGGAAGAACTTC 57

Qy     181 ctagcctgaagatttgtcatagtgtctgctttctagatatctgggaaagatttg 234
           ||||||||||||||||||||||||| ||||||||||||||||  |||| ||
Db      56 CTAGCCTGAAGATTTGTCATAGTGTCTCCTTTCTAGATATCTGGGTCAGATGTG 3


RESULT   4
BF901203/c
LOCUS       BF901203      383 bp    mRNA            EST       18-JAN-2001
DEFINITION  PM4-MT0234-081200-001-e10 MT0234 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF901203
VERSION     BF901203.1  GI:12292662
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-MT0234-
            081200-001-e10&t3=2000-12-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 383.
FEATURES             Location/Qualifiers
     source          1. .383
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0234"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       98 a    107 c     74 g    104 t
ORIGIN


  Query Match             21.0%;  Score 226.6;  DB 170;  Length 383;
  Best Local Similarity   98.3%;  Pred. No. 2.1e-45;
  Matches  229;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ggatgatacaagagccaagaagggacatttgagttgtgtcgcttagataggaaagggatc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 GGATGATACAAGAGCCAAGAAGGGACATTTGAGTTGTGTCGCTTAGATAGGAAAGGGATC 184

Qy      61 cagggaaaatcaacagtaagtgaggatgagcagcgtctcttggttttcattgaggataga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 CAGGGAAAATCAACAGTAAGTGAGGATGAGCAGCGTCTCTTGGTTTTCATTGAGGATAGA 124

Qy     121 gtaagagattgagtttagattgcaacagaaggaattagtttagataccaggaagaacttc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 GTAAGAGATTGAGTTTAGATTGCAACAGAAGGAATTAGTTTAGATACCAGGAAGAACTTC 64

Qy     181 ctagcctgaagatttgtcatagtgtctgctttctagatatctgggaaagattt 233
           ||||||||||||||||||||||||||||||||||||||||||||  ||  ||
Db      63 CTAGCCTGAAGATTTGTCATAGTGTCTGCTTTCTAGATATCTGGGTCAGTGTT 11


RESULT   5
BE980729
LOCUS       BE980729      528 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-BG2-bcm-d-06-0-UI.s1 NIH_BMAP_MSC_S1 Mus musculus cDNA clone
            UI-M-BG2-bcm-d-06-0-UI 3', mRNA sequence.
ACCESSION   BE980729
VERSION     BE980729.1  GI:10649077
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            spinal cord tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BG2-bcm-d-06-0-UI"
                     /clone_lib="NIH_BMAP_MSC_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MSC_S1 library is a subtracted library derived
                     NIH_BMAP_MSC_N. NIH_BMAP_MSC_N was made from  mouse spinal
                     cord tissue. For a detailed description of the library
                     from which this clone was  derived, please visit our web
                     site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_MSC_S1
                     TAG_TISSUE=spinal-cord
                     TAG_SEQ=TCGAA"
```

```
BASE COUNT      115 a    118 c    131 g    164 t
ORIGIN


  Query Match             10.9%;  Score 117.6;  DB 142;  Length 528;
  Best Local Similarity   64.3%;  Pred. No. 1.4e-18;
  Matches  297;  Conservative    0;  Mismatches  139;  Indels   26;  Gaps    7;

Qy     210 tttctagatatctgggaaagatttgataatagttgtttgtgaatagaaaggaggatatga 269
           ||| |  |  ||| ||||||  ||| | | || |||||||    || |||| ||||  |
Db      45 TTTTTTTAAGTCTAGGAAAGTCTTGCTGACAGCTGTTTGTAGGAAGGAAGGTGGAT-CGG 103

Qy     270 tgtttttattggccattttgcgggactcttcgacttcttgctgctgtctcttgaggatac 329
           | |  |   |    | |    | ||| || |||||  |||||||  |||| | ||  ||
Db     104 GGCTGCTGCGGCTGCTCTGTGTGAACTGTTTGACTTGATGCTGCTACCTCTCGTGGTGAC 163

Qy     330 attccaattccatcctggcgagatccaagtgcttacgtactgtctccttagctgccttag 389
           |||   ||  || ||||| |||| || |||||| | | |      ||||||||||||||
Db     164 ATTGTTATAACACCCTGGAGAGACCCGAGTGCTCATGGA--TGTCCCTTAGCTGCCTTAG 221

Qy     390 ------agtaaacgatcatcagttcaatggaccaaaatcaccttcagccatgtggtttct 443
                 |  ||||||   ||  |||||| ||| | || ||||||||||    ||
Db     222 TCGGGAGAACCATCATCATCCACTCGGTGGACCCAAACCGCCATCAGCCATGTG---ACT 278

Qy     444 tcatcatcatggatttcttttggttgacaaacattctggctctcagatgcaaaaagtcac 503
           || | |||||||  ||||||||||||||||||   |||||||||     |   ||||| |||
Db     279 TCCTAATCATGGGTTTCTTTTGGTTGACAA--GTTCTGGCTCAGGTGT---AAAAGCCAC 333

Qy     504 actgggaaatgaactgtaagtggtgaaattagttttggtatttaatttaaaactacattt 563
           ||   ||||||| ||||||| ||||  || | ||| | || ||| ||||||| || || |
Db     334 ACGAAGAAATGAGCTGTAAGCGGTGGCATCACTTTGGATACTTAGTTTAAAAGTATATGT 393

Qy     564 tagtttttctcttctcttctatgttgcaatgaatgtaaagtatttgggatccagtgctta 623
             | |||||| |  ||           ||   ||    ||| ||| ||  | ||    |
Db     394 GTGATTTTCTATGTTC---------CAAAACTATCATGAGTTTTTTGGCCCGAGCAAGTG 444

Qy     624 taaacctttccttcctttgtgcacagaatgtaactagcaagc 665
           || ||||  |||||||||||| |||| ||| |||||  |||
Db     445 TAGACCTCCCCTTCCTTTGTGCCCAGACTGTGACTAGTGAGC 486


RESULT   6
AI340919
LOCUS       AI340919      154 bp    mRNA            EST       18-MAR-1999
DEFINITION  tb52f04.x1 NCI_CGAP_Br15 Homo sapiens cDNA clone IMAGE:2057983 3',
            mRNA sequence.
ACCESSION   AI340919
VERSION     AI340919.1  GI:4077846
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.
            student, Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 254    Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .154
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2057983"
                     /clone_lib="NCI_CGAP_Br15"
                     /sex="female"
                     /tissue_type="adenocarcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     adenocarcinoma tissue, cDNA made by oligo-dT  priming.
                     Directionally cloned.  Size-selected on agarose  gel,
                     average insert size 400 bp.  Primary library,
                     non-amplified."
BASE COUNT       62 a     21 c     33 g     38 t
ORIGIN


  Query Match             10.0%;  Score 108;  DB 19;  Length 154;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-16;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 70

Qy    1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcgtg 1078
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      71 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGTG 118


RESULT   7
BE043150
LOCUS       BE043150      115 bp    mRNA            EST       08-JUN-2000
DEFINITION  hk48f01.y1 NCI_CGAP_Ov34 Homo sapiens cDNA clone IMAGE:2999929 5',
            mRNA sequence.
ACCESSION   BE043150
VERSION     BE043150.1  GI:8360203
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Putative full length read
            The vector to vector length is 116
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2999929"
                     /clone_lib="NCI_CGAP_Ov34"
                     /sex="female"
                     /tissue_type="borderline ovarian carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     borderline ovarian carcinoma, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
```

```
                non-amplified. "
BASE COUNT       37 a     19 c     24 g     35 t
ORIGIN

  Query Match             9.9%;  Score 107;  DB 162;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 4e-16;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 60

Qy    1031 aacattatgctactttcagatattaaaaatggtgttcctttgaatcgt 1077
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGT 107


RESULT   8
AI054382/c
LOCUS       AI054382      144 bp    mRNA            EST       15-JUL-1998
DEFINITION  q164d10.x1 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1861267 3',
            mRNA sequence.
ACCESSION   AI054382
VERSION     AI054382.1  GI:3322169
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 144)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .144
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1861267"
                     /clone_lib="NCI_CGAP_Ov26"
                     /sex="female"
                     /tissue_type="papillary serous carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     papillary serous ovarian carcinoma, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       37 a     31 c     22 g     54 t
ORIGIN

  Query Match             9.9%;  Score 107;  DB 15;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-16;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 66

Qy    1031 aacattatgctactttcagatattaaaaatggtgttcctttgaatcgt 1077
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      65 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGT 19


RESULT   9
AI349657
LOCUS       AI349657      153 bp    mRNA            EST       08-APR-1999
DEFINITION  tb58b02.x1 NCI_CGAP_Br15 Homo sapiens cDNA clone IMAGE:2058507 3',
            mRNA sequence.
ACCESSION   AI349657
VERSION     AI349657.1  GI:4086863
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.;
            student, Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 206   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .153
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2058507"
                     /clone_lib="NCI_CGAP_Br15"
                     /sex="female"
                     /tissue_type="adenocarcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     adenocarcinoma tissue, cDNA made by oligo-dT  priming.
                     Directionally cloned.  Size-selected on agarose  gel,
                     average insert size 400 bp.  Primary library,
                     non-amplified."
BASE COUNT       53 a     25 c     34 g     41 t
ORIGIN

  Query Match             9.9%;  Score 107;  DB 19;  Length 153;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-16;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 81

Qy    1031 aacattatgctactttcagatattaaaaatggtgttcctttgaatcgt 1077
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      82 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGT 128


RESULT  10
AI289554
LOCUS       AI289554      155 bp    mRNA            EST       30-NOV-1998
DEFINITION  qw42c04.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1993734 3',
            mRNA sequence.
```

```
ACCESSION   AI289554
VERSION     AI289554.1  GI:3932734
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1993734"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma, cDNA made by oligo-dT  priming.  Directionally
                     cloned.  Size-selected on agarose  gel, average insert
                     size 500 bp.  Primary library, non-amplified.  "
BASE COUNT       65 a     21 c     31 g     38 t
ORIGIN
```

```
  Query Match             9.9%;  Score 107;  DB 18;  Length 155;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-16;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 70

Qy    1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcgt 1077
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      71 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGT 117
```

```
RESULT  11
AI792441/c
LOCUS       AI792441      173 bp    mRNA            EST       02-JUL-1999
DEFINITION  qi64d10.y5 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1861267 5',
            mRNA sequence.
ACCESSION   AI792441
VERSION     AI792441.1  GI:5340157
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173)
  AUTHORS   NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Dental Research,
            Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: see original entry for original citation
            information
            This 5' resequenced clone has no previous 5' data to verify this
            new read against
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .173
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1861267"
                     /clone_lib="NCI_CGAP_Ov26"
                     /sex="female"
                     /tissue_type="papillary serous carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     papillary serous ovarian carcinoma, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       43 a     37 c     26 g     67 t
ORIGIN
```

```
  Query Match             9.9%;  Score 107;  DB 102;  Length 173;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-16;
  Matches  107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 99

Qy    1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcgt 1077
           |||||||||||||||||||||||||||||||||||||||||||||||
Db      98 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGT 52
```

```
RESULT  12
AI341041
LOCUS       AI341041      163 bp    mRNA            EST       18-MAR-1999
DEFINITION  tb55h06.x1 NCI_CGAP_Br15 Homo sapiens cDNA clone IMAGE:2058299 3',
            mRNA sequence.
ACCESSION   AI341041
VERSION     AI341041.1  GI:4077968
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.
            student, Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
```

```
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 250   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .163
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2058299"
                      /clone_lib="NCI_CGAP_Br15"
                      /sex="female"
                      /tissue_type="adenocarcinoma"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                      adenocarcinoma tissue, cDNA made by oligo-dT  priming.
                      Directionally cloned.  Size-selected on agarose  gel,
                      average insert size 400 bp.  Primary library,
                      non-amplified."
BASE COUNT       62 a      26 c      35 g      40 t
ORIGIN


  Query Match             9.9%;  Score 106.4;  DB 19;  Length 163;
  Best Local Similarity   99.1%;  Pred. No. 6.1e-16;
  Matches  107;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 83

Qy    1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcgtg 1078
           |||||||||||||||||||||||||||||||||||||||||||||| |
Db      84 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCGCG 131


RESULT  13
AI611456
LOCUS       AI611456      141 bp    mRNA          EST       21-APR-1999
DEFINITION  tt61d04.x1 NCI_CGAP_HSC4 Homo sapiens cDNA clone IMAGE:2245255 3',
            mRNA sequence.
ACCESSION   AI611456
VERSION     AI611456.1  GI:4620623
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .141
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2245255"
                      /clone_lib="NCI_CGAP_HSC4"
                      /tissue_type="CD34+, CD38- from normal bone marrow donor"
                      /lab_host="DH10B"
                      /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
                      lymphoid tissue, cDNA made by oligo-dT  priming.
                      Directionally cloned.  Size-selected on agarose  gel,
                      average insert size 500 bp.  Primary library,
                      non-amplified.   cDNA Library Preparation: David B.
                      Krizman, Ph.D. Reference:  Krizman et al. (1996) Cancer
                      Research 56:5380-5383."
BASE COUNT       49 a      21 c      32 g      39 t
ORIGIN


  Query Match             9.8%;  Score 106;  DB 22;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-16;
  Matches  106;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 70

Qy    1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcg 1076
           |||||||||||||||||||||||||||||||||||||||||||||
Db      71 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCG 116


RESULT  14
AI309341
LOCUS       AI309341      147 bp    mRNA          EST       08-APR-1999
DEFINITION  tb51a04.x1 NCI_CGAP_Br15 Homo sapiens cDNA clone IMAGE:2057838 3',
            mRNA sequence.
ACCESSION   AI309341
VERSION     AI309341.1  GI:4004212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.
            student, Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 236   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .147
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2057838"
                      /clone_lib="NCI_CGAP_Br15"
                      /sex="female"
                      /tissue_type="adenocarcinoma"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                      adenocarcinoma tissue, cDNA made by oligo-dT  priming.
                      Directionally cloned.  Size-selected on agarose  gel,
                      average insert size 400 bp.  Primary library,
                      non-amplified."
BASE COUNT       57 a      22 c      31 g      37 t
ORIGIN
```

```
  Query Match             9.8%;  Score 106;  DB 18;  Length 147;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-16;
  Matches  106;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 70

Qy   1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcg 1076
          |||||||||||||||||||||||||||||||||||||||||||||
Db     71 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCG 116


RESULT  15
AI313310
LOCUS       AI313310      150 bp    mRNA            EST       16-FEB-1999
DEFINITION  ta92d09.x2 NCI_CGAP_Brn20 Homo sapiens cDNA clone IMAGE:2051537 3',
            mRNA sequence.
ACCESSION   AI313310
VERSION     AI313310.1  GI:4018915
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Robert Jenkins, M.D., Mark Israel, M.D., Jim
            Jacobson, Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 227   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2051537"
                     /clone_lib="NCI_CGAP_Brn20"
                     /tissue_type="oligodendroglioma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pAMP1; mRNA made from
                     oligodendroglioma tissue, cDNA made by oligo-dT  priming.
                     Directionally cloned.  Size-selected on agarose  gel,
                     average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       60 a     22 c     31 g     37 t
ORIGIN


  Query Match             9.8%;  Score 106;  DB 18;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-16;
  Matches  106;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    971 cagaagggatgcctgctgtaaacaagcagtatgtatggttgtaccaatgcctattggctg 1030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 CAGAAGGGATGCCTGCTGTAAACAAGCAGTATGTATGGTTGTACCAATGCCTATTGGCTG 70

Qy   1031 aacattatgctactttcagatattaaaatggtgttcctttgaatcg 1076
          |||||||||||||||||||||||||||||||||||||||||||||
Db     71 AACATTATGCTACTTTCAGATATTAAAATGGTGTTCCTTTGAATCG 116
```

Search completed: April 30, 2001, 12:17:15
Job time: 12504 sec

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:15:00 ; Search time 22667.1 Seconds
(without alignments)
467.812 Million cell updates/sec

Title: US-09-721-183-18
Perfect score: 1722
Sequence: 1 tgcaccaggatgactctgaa.....acaaaaaaaaaaaaaaaaaa 1722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1722	100.0	1722	88	AF016492	AF016492 Homo sapi
2	1481.2	86.0	1855	93	HUMUDPGTA	J05428 Human 3,4-c
3	1478.6	85.9	2799	93	HSUGT2BIO	X63359 H.sapiens U
4	1457.8	84.7	1753	88	AF016310	AF016310 Macaca fa
5	1444.4	83.9	2648	93	MFU91582	U91582 Macaca fasc
6	1441.2	83.7	1768	88	AF112113	AF112113 Macaca fa
7	1375.6	79.9	2079	92	HSAJ5162	AJ005162 Homo sapi
8	1372.4	79.7	2097	88	AF064200	AF064200 Homo sapi
9	1369.6	79.5	1662	88	AF294901	AF294901 Macaca mu
10	1368	79.4	1662	88	AF294902	AF294902 Macaca mu
11	1366	79.3	2091	88	AF081793	AF081793 Homo sapi

	12	1364.4	79.2	2093	93	HSUDPGT	Y00317 Human mRNA
	13	1351.2	78.5	2108	88	AF112112	AF112112 Macaca fa
	14	1238	71.9	1650	9	AX040087	AX040087 Sequence
	15	1219.2	70.8	2107	93	HSU59209	U59209 Homo sapien
	16	1213.6	70.5	2090	93	HSU08854	U08854 Human UDP g
	17	1212	70.4	2114	88	AF180322	AF180322 Homo sapi
	18	1177.4	68.4	2075	88	AF072223	AF072223 Macaca fa
	19	1175	68.2	2123	93	HSU06641	U06641 Human UDP g
	20	1011	58.7	1671	7	RABUGT2BB	L01082 Oryctolagus
	21	995.2	57.8	2818	7	OCU72742	U72742 Oryctolagus
	22	974.4	56.6	1832	7	RABUGT2BA	L01081 Oryctolagus
	23	961	55.8	1961	94	RATUDPGTP	M13506 Rat liver U
	24	895.4	52.0	1844	95	RNU06273	U06273 Rattus norv
	25	894.2	51.9	1858	94	MMUDPGT	X06358 Mouse mRNA
	26	890	51.7	1716	95	RNUDPGT	Y00156 Rat mRNA fo
	27	887.4	51.5	1846	95	RNU06274	U06274 Rattus norv
	28	885.2	51.4	1819	94	RATGRT	M31109 Rat UDP-glu
	29	865.4	50.3	1858	94	RATUDPA	M33746 Rat UDP glu
	30	856.2	49.7	1947	95	RNU27518	U27518 Rattus norv
	31	853.4	49.6	2634	94	RATUDPGTR	J02589 Rat UDP glu
	32	816.4	47.4	1846	95	RNUDPGTR	X03478 Rat liver m
	33	781.2	45.4	2547	94	AF175221	AF175221 Cavia por
	34	765.2	44.4	1766	92	HSA6054	AJ006054 Homo sapi
	35	741	43.0	1584	94	AF184901	AF184901 Mus muscu
	36	709.8	41.2	1831	95	RRUGTGENE	X57565 Rattus norv
	37	706.4	41.0	2897	7	RABUGT2BC	L01083 Oryctolagus
	38	701.2	40.7	2786	89	AK025587	AK025587 Homo sapi
	39	690.4	40.1	897	88	AF179877	AF179877 Homo Sapi
c	40	678.6	39.4	139015	69	AC025488	AC025488 Homo sapi
	41	677.2	39.3	1021	88	AF179876	AF179876 Homo Sapi
	42	637	37.0	138665	72	AC055794	AC055794 Homo sapi
	43	637	37.0	212904	66	AC021146	AC021146 Homo sapi
	44	587.2	34.1	1007	88	AF179879	AF179879 Homo sapi
c	45	584.8	34.0	27310	88	AF135416	AF135416 Homo sapi

ALIGNMENTS

RESULT 1
AF016492
LOCUS AF016492 1722 bp mRNA PRI 31-JUL-1998
DEFINITION Homo sapiens UDP-glucuronosyltransferase 2B mRNA, complete cds.
ACCESSION AF016492
VERSION AF016492.1 GI:3360272
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1722)
AUTHORS Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
TITLE Isolation and characterization of a human orphan UDP-glucuronosyltransferase, UGT2B11
JOURNAL Biochem. Biophys. Res. Commun. 248 (1), 44-50 (1998)
MEDLINE 98340847
REFERENCE 2 (bases 1 to 1722)
AUTHORS Levesque,E., Beaulieu,M., Hum,D.W. and Belanger,A.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1997) Molecular Endocrinology, Laval University, 2705 Laurier Blvd., Ste-Foy, Quebec G1V 4G2, Canada
FEATURES Location/Qualifiers
source 1. .1722
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 10. .1599
/note="LN117"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B"
/protein_id="AAC27891.1"
/db_xref="GI:3360273"
/translation="MTLKWTSVLLLIHLSCYFSSGSCGKVLVWAAEYSHWMNMKTILK
ELVQRGHEVTVLASSASILFDPNDASTLKFEVYPTSLTKTEFENIIMQQVKRWSDIRK
DSFWLYFSQEQEILWELYDIFRNFCKDVVSNKKVMKKLQESRFDIVFADAVFPCGELL
AALLNIRFVYSLRFTPGYTIERHSGGLIFPPSYIPIVMSKLSDQMTFMERVKNMIYVL
YFDFWFQMSDMKKWDQFYSEVLGRPTTLFETMGKADIWLMRNSWSFQFPHPFLPNVDF
VGGFHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSVISNMTAERANVIATALAKIPQK
VLWRFDGNKPDALGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
IPLFFDQPDNIAHMKAKGAAVRLDFNTMSSTDLLNALKTVINDPLYKENIMKLSRIQH
DQPVKPLDRAVFWIEFVMPHKGAKHLRVAAHDLTWFQYHSLDVIGFLLACVATVIFII
TKFCLFCFWKFARKGKKGKRD"
BASE COUNT 523 a 319 c 371 g 509 t
ORIGIN

Query Match 100.0%; Score 1722; DB 88; Length 1722;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
Db       1 TGCACCAGGATGACTCTGAAATGGACTTCAGTTCTTCTGCTGATACATCTCAGTTGTTAC 60

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
Db      61 TTTAGCTCTGGGAGTTGTGGAAAAGTGCTGGTGTGGGCCGCAGAATACAGCCATTGGATG 120

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
Db     121 AATATGAAGACAATCCTGAAAGAGCTTGTTCAGAGAGGTCATGAGGTGACTGTACTGGCA 180

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
Db     181 TCTTCAGCTTCCATTCTTTTTGATCCCAATGATGCATCCACTCTTAAATTTGAAGTTTAT 240

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
Db     241 CCTACATCTTTAACTAAAACTGAATTTGAGAATATCATCATGCAACAGGTTAAGAGATGG 300

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
Db     301 TCAGACATTCGAAAAGATAGCTTTTGGTTATATTTTTCACAAGAACAAGAAATCCTGTGG 360

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db     361 GAATTATATGACATATTTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAAGTTATG 420

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db     421 AAAAAACTACAAGAGTCAAGATTTGACATCGTTTTTGCAGATGCTGTTTTTCCCTGTGGT 480

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db     481 GAGCTGCTGGCTGCGCTACTTAACATACGGTTTGTGTACAGTCTCCGCTTTACTCCTGGC 540

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db     541 TACACAATTGAAAGGCACAGTGGAGGACTGATTTTCCCTCCTTCCTACATACCTATTGTT 600

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db     601 ATGTCAAAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTG 660

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db     661 CTTTATTTTGACTTTTGGTTCCAAATGTCTGATATGAAGAAGTGGGATCAGTTTTACAGT 720

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db     721 GAAGTTTTAGGAAGACCCACTACCTTATTTGAGACAATGGGAAAAGCTGACATATGGCTT 780

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db     781 ATGCGAAACTCCTGGAGTTTTCAATTTCCTCATCCATTCTTACCAAACGTTGATTTTGTT 840

Qy     841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
```

```
Db    841 GGAGGATTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTACAG 900

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAGTGATAAGTAACATGACA 960

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCAGAAAGGGCCAATGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1020

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGATTTGACGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGGCTGTACAAGTGGATA 1080

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1140

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttttt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGGTGGGCATTCCATTGTTTTTT 1200

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GATCAACCTGATAACATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGATTGGACTTC 1260

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AACACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTTTA 1320

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TATAAAGAGAATATTATGAAATTATCAAGAATTCAACATGATCAACCAGTAAAGCCCCTG 1380

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCCCCACAAAGGAGCCAAACACCTTCGA 1440

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GTTGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTTCTG 1500

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTGGCCTGTGTGGCAACTGTGATATTTATCATCACAAAGTTTTGTCTGTTTTGTTTCTGG 1560

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AAGTTTGCTAGAAAAGGGAAGAAGGGAAAAAGAGATTAGTTATGTCTGACATTTGAAGCT 1620

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GGAAAACCAGATAGATAGGACAACTTCAGTTTATTCCAGCAAGAAAGAAAAGATTGTTAT 1680

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
          ||||||||||||||||||||||||||||||||||||||||
Db   1681 GCAAGATTTCTTTCTTCCTGTGACAAAAAAAAAAAAAAAAAA 1722


RESULT   2
HUMUDPGTA
LOCUS       HUMUDPGTA    1855 bp    mRNA            PRI       03-AUG-1993
DEFINITION  Human 3,4-catechol estrogen UDP-glucuronosyltransferase mRNA,
            complete cds.
ACCESSION   J05428
VERSION     J05428.1  GI:340079
KEYWORDS    3,4-catechol estrogen UDP-glucuronosyltransferase.
SOURCE      Human liver, cDNA to mRNA, clone 63-11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1855)
  AUTHORS   Ritter,J.K., Sheen,Y.Y. and Owens,I.S.
  TITLE     Cloning and expression of human liver UDP-glucuronosyltransferase
            in COS-1 cells: 3,4-catechol estrogens and estriol as primary
            substrates
  JOURNAL   J. Biol. Chem. 265, 7900-7906 (1990)
  MEDLINE   90243659
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by I.S.Owens, 22-FEB-1990.
FEATURES             Location/Qualifiers
     source          1. .1855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             15. .1604
                     /note="UDP-glucuronosyltransferase (EC 2.4.1.17)"
                     /codon_start=1
                     /protein_id="AAA36793.1"
                     /db_xref="GI:340080"
                     /translation="MSVKWTSVILLIQLSFCFSSGNCGKVLVWAAEYSHWMNIKTILD
                     ELIQRGHEVTVLASSASILFDPNNSSALKIEIYPTSLTKTELENFIMQQIKRWSDLPK
                     DTFWLYFSQVQEIMSIFGDITRKFCKDVVSNKKFMKKVQESRFDVIFADAIFPCSELL
                     AELFNIPFVYSLSFSPGYTFEKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVL
                     YFDFWFEIFDMKKWDQFYSEVLGRPTTLSETMGKADVWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEDFVQSSGENGVVVFSLGSMVSNMTEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     IPLFADQPDNIAHMKARGAAVRVDFNTMSSTDLLNALKRVINDPSYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLVCVATVIFIV
                     TKCCLFCFWKFARKAKKGKND"
BASE COUNT      571 a    337 c    394 g    553 t
ORIGIN


  Query Match           86.0%;  Score 1481.2;  DB 93;  Length 1855;
  Best Local Similarity  91.7%;  Pred. No. 0;
  Matches 1579;  Conservative    0;  Mismatches  138;  Indels    5;  Gaps    1;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||||||||||| || ||||||||||||||||  || |||| ||||| || || | || |
Db      6 TGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTTTGCTAATACAACTGAGCTTTTGC 65

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          |||||||||||| |||||||| |||||||||||||| ||||||||||||||||||||||
Db     66 TTTAGCTCTGGGAATTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATACAGCCATTGGATG 125

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          ||||| |||||||||||| | |||||| ||||||||||||||||||||||||||||||||
Db    126 AATATAAAGACAATCCTGGATGAGCTTATTCAGAGAGGTCATGAGGTGACTGTACTGGCA 185

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          |||||||||||||||||||||||||||||  | ||||| |||||||| ||||| |||||
Db    186 TCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAATTTAT 245

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          || |||||||||||||||||| || |||||| |||||||||||||| ||||||||||||
Db    246 CCCACATCTTTAACTAAAACTGAGTTGGAGAATTTCATCATGCAACAGATTAAGAGATGG 305

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          |||||| ||| |||||||| ||||||||||||||||||||| ||| |||||| |||
Db    306 TCAGACCTTCCAAAAGATACATTTTGGTTATATTTTTCACAAGTACAGGAAATCATGTCA 365

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
            |||   |||||||  |||||| ||||||||||||||||||||||||||||| |||||
Db    366 ATATTTGGTGACATAACTAGAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAAATTTATG 425

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          |||||| |||||||||||||||||| || |||||||||||||| ||||||||||| ||
Db    426 AAAAAAGTACAAGAGTCAAGATTTGACGTCATTTTTGCAGATGCTATTTTTCCCTGTAGT 485

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||| |||| |||||||||  |||||||||||||||| ||||  ||||||||
Db    486 GAGCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCTCAGCTTCTCTCCTGGC 545

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          |||||  |||||| ||| |||||||||  | |||||||||||||||||| ||||| |||||
```

```
Db    546 TACACTTTTGAAAAGCATAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTGTTGTT 605

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 ATGTCAGAATTAACTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTG 665

Qy    661 ctttatttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          ||||| |||||||||||||||| |||| | ||| ||||||||||||||||||||||| |||
Db    666 CTTTACTTTGACTTTTGGTTCGAAATATTTGACATGAAGAAGTGGGATCAGTTTTATAGT 725

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||||| |||| |||||||||||| ||||||||| ||||||||
Db    726 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTATGGCTT 785

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || |||||||||||||| |||||| |||||||||||| |||||||||| ||||||||||||
Db    786 ATTCGAAACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATTTTGTT 845

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||| |||||||||||||| || |||||||||
Db    846 GGAGGACTCCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAG 905

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          ||||||||||||||||||||||||||||||||||||||||||| || | ||||||||||||
Db    906 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACA 965

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||| |||||||| ||||||| ||| ||||||||||||||||||||||||
Db    966 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGG 1025

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||| ||||||||||||||||||||||||| |||||||||
Db   1026 AGATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATA 1085

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   1086 CCCCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCC 1145

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttttt 1200
          |||||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1146 AATGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCC 1205

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          |||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||||
Db   1206 GATCAACCTGATAACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGGACTTC 1265

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          |||||||||||||||||| |||||||||| |||||| ||||||||||||||||||| |
Db   1266 AACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATCCTTCA 1325

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          |||||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
Db   1326 TATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTG 1385

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db   1386 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGG 1445

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1446 GTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTG 1505

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||| |||||||||||||||||||||||||||||| ||||||| | ||||||||||||||||
Db   1506 CTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTTTCTGG 1565

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          |||||||||||||||  ||||||||||||  ||||||||| ||||| ||||||||||
Db   1566 AAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTATATCTGAGATTTGAAGCT 1625

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          |||||||| ||||| |  ||| |||||||||||||||||||     ||||||||| ||    (
Db   1626 GGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAAG-----AAAGATTGTGAT 1680
```

```
Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
          |||||||||||||||||||| |||||||||||||||| |||
Db   1681 GCAAGATTTCTTTCTTCCTGAGACAAAAAAAAAAAAAAGAAA 1722


RESULT   3
HSUGT2BIO
LOCUS       HSUGT2BIO    2799 bp    mRNA            PRI       19-JUL-1994
DEFINITION  H.sapiens UGT2BIO mRNA for udp glucuronosyltransferase.
ACCESSION   X63359
VERSION     X63359.1  GI:516149
KEYWORDS    UDP-glucuronosyltransferase; ugt2bio gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Mackenzie,P.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1991) P.I. Mackenzie, Flinders University, Dept
            of Clinical Pharmacology, School of Medicine, Flinders Medical
            Centre, Bedford Park SA 5042, AUSTRALIA
REFERENCE   2  (bases 1 to 2799)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
  MEDLINE   93326164
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11"
                     /clone="UGTh-46"
     gene            11. .1597
                     /gene="UGT2 BIO"
     CDS             11. .1597
                     /gene="UGT2 BIO"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="CAA44961.1"
                     /db_xref="GI:516150"
                     /db_xref="SWISS-PROT:P36537"
                     /translation="MALKWTTVLLIQLSFYFSSGSCGKVLVWAAEYSLWMNMKTILKE
                     LVQRGHEVTVLASSASILFDPNDSSTLKLEVYPTSLTKTEFENIIMQLVKRLSEIQKD
                     TFWLPFSQEQEILWAINDIIRNFCKDVVSNKKLMKKLQESRFDIVFADAYLPCGELLA
                     ELFNIPFVYSHSFSPGYSFERHSGGFIFPPSYVPVVMSKLSDQMTFMERVKNMLYVLY
                     FDFWFQIFNMKKWDQFYSEVLGRPTTLSETMRKADIWLMRNSWNFKFPHPFLPNVDFV
                     GGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNMTEERANVIATALAKIPQKV
                     LWRFDGNKPDALGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGI
                     PLFFDQPDNIAHMKAKGAAVRVDFNTMSSTDLLNALKTVINDPSYKENIMKLSRIQHD
                     QPVKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWFQYHSLDVIGFLLACVATVLFIIT
                     KCCLFCFWKFARKGKKGKRD"
     polyA_signal    2772. .2788
BASE COUNT      792 a    605 c    569 g    833 t
ORIGIN
```

```
  Query Match           85.9%;  Score 1478.6;  DB 93;  Length 2799;
  Best Local Similarity   92.0%;  Pred. No. 0;
  Matches 1572;  Conservative    0;  Mismatches  134;  Indels    3;  Gaps    1;

Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
            ||||| |||||| |||||||||||||| |||   |||||||||||||| ||||||| ||||
Db       2 TGCACAAGGATGGCTCTGAAATGGACTACAG---TTCTGCTGATACAACTCAGTTTTTAC 58

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
           |||||||||||||||||||||||| |||||||| |||||||||||||||||| |||||||
Db      59 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTATGGGCCGCAGAATACAGCCTTTGGATG 118
```

```
Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
Db    119 AATATGAAGACAATCCTGAAAGAACTTGTTCAGAGAGGTCATGAGGTGACTGTACTGGCA 178

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
Db    179 TCTTCAGCTTCCATTCTTTTTGATCCCAACGACTCATCCACTCTTAAACTTGAAGTTTAT 238

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
Db    239 CCTACATCTTTAACTAAAACTGAATTTGAGAATATCATCATGCAATTGGTTAAGAGATTG 298

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
Db    299 TCAGAAATTCAAAAAGATACATTTTGGTTACCTTTTTCACAAGAACAAGAAATCCTGTGG 358

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db    359 GCAATTAATGACATAATTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAACTTATG 418

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db    419 AAAAAACTACAAGAGTCAAGATTTGACATCGTTTTTGCAGATGCTTATTTACCCTGTGGT 478

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db    479 GAGCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCACAGCTTCAGTCCTGGC 538

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db    539 TACTCATTTGAAAGGCACAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTGTTGTT 598

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db    599 ATGTCAAAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGCTCTATGTG 658

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db    659 CTTTATTTTGACTTTTGGTTCCAAATATTTAATATGAAGAAGTGGGATCAGTTTTACAGT 718

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db    719 GAAGTTTTAGGAAGACCCACTACATTATCTGAGACAATGAGGAAAGCTGACATATGGCTT 778

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db    779 ATGCGAAACTCCTGGAATTTTAAATTTCCTCATCCATTCTTACCAAATGTTGATTTTGTT 838

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
Db    839 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTACAG 898

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db    899 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACA 958

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db    959 GAAGAAAGGGCCAACGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGTTCTTTGG 1018

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db   1019 AGATTTGATGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGACTGTACAAGTGGATA 1078

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db   1079 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1138

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttttt 1200
Db   1139 AATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGGTGGGCATTCCATTGTTTTTT 1198

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db   1199 GATCAACCTGATAATATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGAGTGGACTTC 1258

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db   1259 AACACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTTCA 1318

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db   1319 TATAAAGAGAATATTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTG 1378

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
Db   1379 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCCAAACATCTTCGA 1438

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db   1439 GTTGCAGCCCACAACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTG 1498

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
Db   1499 CTGGCTTGTGTGGCAACCGTGCTATTTATCATCACAAAGTGTTGTCTGTTTTGTTTCTGG 1558

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
Db   1559 AAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGGGATTAGTTATATCTGAGATTTGAAGCT 1618

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
Db   1619 GGGGAATTCCGTTTATTGAAGATTCAGGTTAACCTGAATCAAGTTAACCCAGTCTCAAAT 1678

Qy   1681 gcaagatttctttcttcctgtgacaaaaa 1709
Db   1679 GCTCACTTATCCTATCTCCTTGGCACAAA 1707
```

```
RESULT   4
AF016310
LOCUS       AF016310     1753 bp    mRNA            PRI       30-DEC-1998
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase (UGT2B18) mRNA,
            complete cds.
ACCESSION   AF016310
VERSION     AF016310.1  GI:4079706
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1753)
  AUTHORS   Beaulieu,M., Levesque,E., Barbier,O., Turgeon,D., Belanger,G.,
            Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Molecular Endocrinology, Laval University,
            2705 Laurier Blvd, Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .1753
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     gene            1. .1753
                     /gene="UGT2B18"
     CDS             1. .1590
                     /gene="UGT2B18"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAC98726.1"
                     /db_xref="GI:4079707"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENIIRQQIKRWSELPK
                     DTFWLYFSQMQEIMWKFGDITRNFCKDVVSNKKLMKKLQKSRFDVVFADAIFPCSELL
                     AELLNTPLVYSLRFTPGYNFEKHCGGFLFPPSYVPVVMSELSDHMTFMERVKNMIYML
                     YFDFCFQIYAMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
```

VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVTNMKEERANVIASALAQIPQK
VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGSNGIYEAIYHGVPMVG
IPLFADQPDNIAHMKAKGAAVRLDFDTMSSTDLVNALKTVINDPLYKENVMKLSRIQH
DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFII
MKCCLFCFWKFARKGKKGKSD"
BASE COUNT 549 a 323 c 362 g 519 t
ORIGIN

Query Match 84.7%; Score 1457.8; DB 88; Length 1753;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1561; Conservative 0; Mismatches 147; Indels 5; Gaps 1;

```
Qy      10 atgactctgaaatggacttcagttcttctgctgatacatctcagttgttactttagctct 69
           ||| || ||||||||||||||||  ||||||| ||||| || || | ||||||||||||||
Db       1 ATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACTGAGCTTTTACTTTAGCTCT 60

Qy      70 gggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaag 129
           ||||||||||||||| |||||||||||| |||||||| ||||||||||||||||||||||
Db      61 GGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATGAATATGAAG 120

Qy     130 acaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagct 189
           |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     121 ACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTACTGGCATCTTCAGCT 180

Qy     190 tccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatct 249
           |||||||||||||||||||  | ||||| |||||||| |||||||| ||||||||||
Db     181 TCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTTTTTCCTACATCT 240

Qy     250 ttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacatt 309
           |||||||||||||| |||||||||||||||| ||||||  ||||||||||||||||  ||
Db     241 TTAACTAAAACTGAGTTTGAGAATATCATCAGGCAACAAATTAAGAGATGGTCAGAACTT 300

Qy     310 cgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgggaattatat 369
           | ||||||||  |||||||||||||||||||||   ||||||||| ||||| | ||   |
Db     301 CCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAATCATGTGGAAGTTTGGT 360

Qy     370 gacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaacta 429
           ||||||  |||||| |||||||||||||||||||||||||||| | |||||  ||||||
Db     361 GACATAACTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATGAAGAAACTA 420

Qy     430 caagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgctg 489
           ||| |||||||||||||  |||||||||||||||  ||||||||||  ||||| |||||
Db     421 CAAAAGTCAAGATTTGACGTCGTTTTTGCAGATGCTATTTTTCCCTGTAGTGAGTTGCTG 480

Qy     490 gctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggctacacaatt 549
           |||| |||||||||||  ||   ||||||||||||||||||  ||||||||||||   ||
Db     481 GCTGAGCTACTTAACACACCCCTTGTGTACAGTCTCCGCTTCACTCCTGGCTACAATTTT 540

Qy     550 gaaaggcacagtggaggactgattttccctccttcctacatacctattgttatgtcaaaa 609
           |||| |||  |||||||| |  |||||||||| |||||  ||||| |||||||||||| ||
Db     541 GAAAAGCATTGTGGAGGATTTCTTTTCCCTCCCTCCTATGTACCTGTTGTTATGTCAGAA 600

Qy     610 ttaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtgctttatttt 669
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTAAGTGATCACATGACTTTCATGGAGAGGGTAAAAAATATGATCTATATGCTTTATTTT 660

Qy     670 gacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagtgaagtttta 729
           |||||||| |||||||| | ||  |||||||||||||||||||||| ||||||||| ||
Db     661 GACTTTTGTTTCCAAATATATGCCATGAAGAAGTGGGATCAGTTTTATAGTGAAGTTCTA 720

Qy     730 ggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaac 789
           |||||||||||||| |||| ||||||||||| ||||||||||||||||||||| ||||||
Db     721 GGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACATATGGCTTATTCGAAAC 780

Qy     790 tcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggattc 849
           ||||||| |||||| ||||| |||||| |||||||||| ||||||||||||||||||  ||
Db     781 TCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTGTTGGAGGACTC 840

Qy     850 cactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctctgga 909
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     841 CACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTTGTACAGAGCTCTGGA 900

Qy     910 gaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagg 969
           ||||||||||||||||||||||||||||||||||||||| || | | |||||||| || |||||||
Db     901 GAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCACTAACATGAAAGAAGAAAGG 960

Qy     970 gccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgac 1029
           |||||  |||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db     961 GCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTGAT 1020

Qy    1030 gggaataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccagaat 1089
           || || |||||||| ||||||||||||||||||||||| || ||||||||||||||||||
Db    1021 GGAAAAAAACCAGATACCTTAGGTCTCAATACTCGGCTCTATAAGTGGATACCCCAGAAT 1080

Qy    1090 gaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatc 1149
           || |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1081 GATCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGATCCAATGGCATC 1140

Qy    1150 tatgaggcaatctaccatgggatccctatggtgggcattccattgtttttgatcaacct 1209
           || |||||||||||||||| |||||||||||||||||||||||||||  |||||||||
Db    1141 TACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCATTGTTTGCCGATCAACCT 1200

Qy    1210 gataacattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatg 1269
           |||||||||||||||||| ||||| |||||||||||||||||||||| ||||||||
Db    1201 GATAACATTGCTCACATGAAAGCCAAAGGAGCAGCTGTTAGATTGGACTTCGACACAATG 1260

Qy    1270 tcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctttatataaagag 1329
           |||||||||||| || ||||||| |||||||||||||||||||||||||||||||||||||
Db    1261 TCGAGTACAGACTTGGTGAATGCCCTGAAGACAGTAATTAATGATCCTTTATATAAAGAG 1320

Qy    1330 aatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgagca 1389
           ||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    1321 AATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTGGATCGAGCA 1380

Qy    1390 gtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcagcc 1449
           |||||||||||||||||||||||||| |||||||||||  ||||||||||   |||||||
Db    1381 GTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGGCCTGCAGCC 1440

Qy    1450 catgacctcacctggttccagtaccactctttggatgtgattgggtttctgctggcctgt 1509
           |||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db    1441 CATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTGGGTTCCTGCTGGCCTGT 1500

Qy    1510 gtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctggaagtttgct 1569
           |||||||||||||||||||||||||  || | ||||||||||||||||||||||||||||
Db    1501 GTGGCAACTGTGATATTTATCATCATGAAATGTTGTCTGTTTTGTTTCTGGAAGTTTGCT 1560

Qy    1570 agaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagctggaaaacca 1629
           |||||||| |||||||||||||| |||||||||| ||||| ||||||||||||||||||
Db    1561 AGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTGAGATTTGAAGCTGGAAAACCT 1620

Qy    1630 gatagataggacaacttcagtttattccagcaagaaagaaaagattgttatgcaagattt 1689
           |||||||  ||| |||||||||||||||||||||||     |||| || | || ||||||||
Db    1621 GATAGATGAGACTACTTCAGTTTATTCCAGCAAG-----AAAGGTTATGATACAAGATTT 1675

Qy    1690 ctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
           |||||||||||||| |||| ||||| ||||||
Db    1676 CTTTCTTCCTGTGACCAAAATAAAAATAAAAAA 1708
```

RESULT 5
MFU91582
LOCUS MFU91582 2648 bp mRNA PRI 25-MAR-1997
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION U91582
VERSION U91582.1 GI:1906053
KEYWORDS .
SOURCE crab-eating macaque.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 2648)
AUTHORS Belanger,G., Beaulieu,M., Levesque,E., Hum,W.H. and Belanger,A.
TITLE Expression and characterization of a novel
UDP-Glucuronosyltransferase, UGT2B9, from cynomolgus monkey

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2648)
  AUTHORS   Belanger,G., Beaulieu,M., Levesque,E., Hum,W.H. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1997) Endocrinology, CHUL, 2705 Boulevard
            Laurier, Ste Foy, QC G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2648
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             28. .1617
                     /note="UGT2B9"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAB50249.1"
                     /db_xref="GI:1906054"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENISMQEVKRWIELPK
                     DTFWLYFSQMQEIMWRFGDIIRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYIFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     SFDFYFQMYDMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVTNMEEERANVIASALAQIPQK
                     VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTKGAAVRLDFDTMSSTDLANRLKTVINDPLYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFVI
                     MKCCLFCFWKFARKGKKGKSD"
BASE COUNT      821 a    499 c    562 g    766 t
ORIGIN
```

```
  Query Match             83.9%;  Score 1444.4;  DB 93;  Length 2648;
  Best Local Similarity   90.4%;  Pred. No. 0;
  Matches 1556;  Conservative    0;  Mismatches  161;  Indels    5;  Gaps    1;

Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
           ||||||||||||  ||  ||||||||||||||||  |||||||  |||||  ||  ||  |  ||||
Db      19 TGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACTGAGCTTTTAC 78

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
           ||||||||||||||||||||||  ||||||||||||||  ||||||||  ||||||||||||
Db      79 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATG 138

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
           ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     139 AATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTCCTGGCA 198

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
           ||||||||||||||||||||||||||||  |  |||||  |||||||  |||||||  | |
Db     199 TCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTCTTT 258

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
           ||||||||||||||||||||  ||||||||||  |||||||  |||||||||||||||||
Db     259 CCTACATCTTTAACTAAAACTGAGTTTGAGAATATCTCCATGCAAGAGGTTAAGAGATGG 318

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
             |||  |||  ||||||||  ||||||||||||||||||||  |||||||||  |||||
Db     319 ATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAATCATGTGG 378

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaatagaaagttatg 420
              ||   |||||||  |||||||  ||||||||||||||||||||||||||||||  | |||
Db     379 AGGTTTGGTGACATAATTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATG 438

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
           ||  |||  ||||||||||||||||||  ||||||||||||||  ||  ||||||||||  ||
Db     439 AAGAAATTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTCCCTGTAGT 498

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
           ||||||||||||  ||||  |||||||||  ||||||||||||||||||||  |||||||||
Db     499 GAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCACTCCTGGC 558

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
           ||||   ||||||  |||  |||||||  |  ||||||||||  ||||||  |||||  ||||
Db     559 TACATTTTTGAAAAGCATTGTGGAGGTTTTCTTTTCCCTCCCTCCTACGTACCTGTTGTT 618
```

```
Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaaatatgatctatgtg 660
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db     619 ATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATATG 678

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
           ||||  ||||||||||||  ||  |||||||  |||  ||||||||||||||||||  |||||  |||
Db     679 CTTTCTTTTGACTTTTATTTTCAAATGTATGACATGAAGAAGTGGGATCAATTTTATAGT 738

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
           ||||||  ||||||||||||||||||  ||||  ||||||||||||  ||||||||||||||||||
Db     739 GAAGTTCTAGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACATATGGCTT 798

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
           ||  ||||||||||||  ||||||  |||||  ||||||  ||||||||||  ||||||||||||
Db     799 ATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTGTT 858

Qy     841 ggaggattccactgcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 900
           ||||||  ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db     859 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTTGTACAG 918

Qy     901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||  ||  |  |||||||  |
Db     919 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCACTAACATGGAA 978

Qy     961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
           |  ||||||||||||  |||||||||  ||||||||||  ||||||||||||||||||||||||
Db     979 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTGTGG 1038

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
           ||||||||  ||  ||||||  |||  |||||||  ||||||||||||||  ||  |||||||||
Db    1039 AGATTTGATGGAAAAAAACCGGATACCTTAGGCCTCAATACTCGGCTCTATAAGTGGATA 1098

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1099 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGAGCC 1158

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
           ||||||||||||||  ||  ||||||||||||  ||||||||||||||||||||||||||||
Db    1159 AATGGCATCTATGAAGCGATCTACCATGGGGTCCCTATGGTGGGCATTCCATTGTTTGCC 1218

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
           |||||  ||||||||||||||||||||||||  |||||  ||||||||||||||||||||||||
Db    1219 GATCAGCCTGATAACATTGCTCACATGAAGACCAAAGGAGCAGCTGTTAGATTGGACTTC 1278

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
           ||||||||||||||||||||||  ||  ||||  ||||||||||||||||||||||||||||||
Db    1279 GACACAATGTCGAGTACAGACTTGGCGAATCGACTGAAGACAGTAATTAATGATCCTTTA 1338

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
           ||||||||||||  ||||||||||||||||||||||||||||||||||||||||  ||||||||
Db    1339 TATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTG 1398

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
           ||||||||||||||||||||||||||||||||||||||  ||||||||||||  ||||||||||
Db    1399 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGG 1458

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db    1459 CCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTGGGTTCCTG 1518

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
           ||||||||||||||||||||||||||  |  ||||  ||  |  ||||||||||||||||||||
Db    1519 CTGGCCTGTGTGGCAACTGTGATATTTGTTATCATGAAATGTTGTCTGTTTTGTTTCTGG 1578

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
           ||||||||||||||||  ||||||||||||||  ||||||||||  |||||  ||||||||||
Db    1579 AAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTGAGATTTGAAGCT 1638

Qy    1621 ggaaaaccagatagatagaacaacttcagtttattccagcaagaaagaaaagattgttat 1680
           ||||||||  |||||||  |||  |||||||  |||||||||  |||     ||||  ||  |  ||
Db    1639 GGAAAACCTGATAGATGAGACTACTTCAGCTTATTCCAGGAAG-----AAAGGTTATGAT 1693
```

```
Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
          |||||||||||||||||||||||| |||||| ||||| | ||
Db   1694 GCAAGATTTCTTTCTTCCTGTGACCAAAAAATAAAAATATAA 1735


RESULT   6
AF112113
LOCUS       AF112113     1768 bp    mRNA            PRI       02-DEC-1999
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase 2B23 precursor,
            mRNA, complete cds.
ACCESSION   AF112113
VERSION     AF112113.1  GI:6502548
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1768)
  AUTHORS   Barbier,O., Levesque,E., Belanger,A. and Hum,D.W.
  TITLE     UGT2B23, a novel uridine diphosphate-glucuronosyltransferase enzyme
            expressed in steroid target tissues that conjugates androgen and
            estrogen metabolites
  JOURNAL   Endocrinology 140 (12), 5538-5548 (1999)
  MEDLINE   20043918
   PUBMED   10579317
REFERENCE   2  (bases 1 to 1768)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .1768
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             28. .1617
                     /EC_number="2.4.1.17"
                     /note="UGT2B23"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B23 precursor"
                     /protein_id="AAF14353.1"
                     /db_xref="GI:6502549"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTALASSASILFDPNNSSALKIEVFPTSLPKPEFENIVTQEIKRWIELPK
                     DTFWLYFSQMQEIMWKFGDIFRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYVFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     YFDFCFQIYDMKKWDQFYTEVLGRHTTLSEIMGKADIWLIRNSWNFQFPHPLLPNVDF
                     IGGLLCKPAKPLPKEMEEFVQSSGENGVVVFTLGSMITNMKEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGVNTRLYKWIPQNDLLGHPKTKAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTRGAAVQLDFDTMSSTDLVNALKTVINDPLYKENVMKLSRIQR
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSFDVIGFLLACVATVIFII
                     MKCCLFCFWKFARKGKKGKSD"
BASE COUNT      546 a    327 c    374 g    521 t
ORIGIN


  Query Match           83.7%;  Score 1441.2;  DB 88;  Length 1768;
  Best Local Similarity 90.2%;  Pred. No. 0;
  Matches 1554;  Conservative    0;  Mismatches  163;  Indels    5;  Gaps    1;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||||||||||| || ||||||||||||||||  ||||||| ||||| || || | ||||
Db     19 TGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACTGAGCTTTTAC 78

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          ||||||||||||||||||||||| ||||||||||||||| |||||||| ||||||||||||
Db     79 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATG 138

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          |||||||||||||||||| |||||||||| ||||||||||||||||||||||| |||||||
Db    139 AATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGCACTGGCA 198

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          |||||||||||||||||||||||| |  ||||| |||||||| |||||||| |
Db    199 TCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTTTTT 258

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          |||||||||||| ||||| |||| |||||||||||  ||| |||| || ||||||||||||
Db    259 CCTACATCTTTACCTAAACCTGAGTTTGAGAATATAGTCACGCAAGAGATTAAGAGATGG 318

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          |||  ||| |||||||| ||||||||||||||||||||   |||||||||| ||||
Db    319 ATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAATCATGTGG 378

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
          | ||  ||||||||||||||||||||||||||||||||||||||||||||||||| | |||
Db    379 AAGTTTGGTGACATATTTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATG 438

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          || |||||||||||||||||||||||| ||||||||||||| || |||||||||| ||
Db    439 AAGAAACTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTCCCTGTAGT 498

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||| |||| |||||||||  ||||||||||||||||||| |||||||||
Db    499 GAGCTGCTGGCTGAGCTATTTAACATACCACTTGTGTACAGTCTCCGCTTCACTCCTGGC 558

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||      |||||| ||| |||||||| |  |||||||||| |||||| ||||| |||||
Db    559 TATGTTTTTGAAAAGCATTGTGGAGGATTTCTTTTCCCTCCCTCCTACGTACCTGTTGTT 618

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 ATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATATG 678

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          |||||||||||||||| |||||||| | ||| |||||||||||||||||||||||| | |
Db    679 CTTTATTTTGACTTTTGTTTCCAAATATATGACATGAAGAAGTGGGATCAGTTTTATACT 738

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||| |||||| |||| ||||| |||||| ||||||||||||||||||
Db    739 GAAGTTCTAGGAAGACACACTACATTATCTGAGATAATGGGGAAAGCTGACATATGGCTT 798

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || |||||||||||| |||||| ||||| |||||| |||||||||| ||||||||| ||
Db    799 ATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTATT 858

Qy    841 ggaggattccactgcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 900
          |||||| ||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    859 GGAGGACTCCTCTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTTGTACAG 918

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          |||||||||||||| |||||||||||||||||||||||||||| | ||||||||| |
Db    919 AGCTCTGGAGAAAACGGTGTTGTGGTGTTTACTCTGGGGTCAATGATCACTAACATGAAA 978

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||| |||||||| |||||||||| |||| ||||||||||||||||||
Db    979 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATTCCACAAAAGGTTCTGTGG 1038

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||| |||||||| |||||||||||| || |||||||||
Db   1039 AGATTTGATGGGAATAAACCAGATACCTTAGGTGTCAATACTCGGCTCTATAAGTGGATA 1098

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   1099 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAAAGCTTTTATAACTCACGGTGGAGCC 1158

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
          |||||||||| |||||||||||||||||| ||||||||||||||||||| |||||
Db   1159 AATGGCATCTACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCACTGTTTGCC 1218

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          |||||||||||||||||||||||||||||| |||  |||||||||||  ||||||||||
Db   1219 GATCAACCTGATAACATTGCTCACATGAAAACCAGAGGAGCAGCTGTTCAATTGGACTTC 1278

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          |||||||||||||||||||| || ||||||||||||||||||||||||||||||||||
```

```
Db   1279 GACACAATGTCGAGTACAGACTTGGTGAATGCACTGAAGACAGTAATTAATGATCCTTTA 1338

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          |||||||||||| |||||||||||||||||||||||| |||||| ||||| |||||||||
Db   1339 TATAAAGAGAATGTTATGAAATTATCAAGAATTCAACGTGATCAGCCAGTGAAGCCCCTG 1398

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          |||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db   1399 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGG 1458

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
            |||||||||||||||||||||||||||||| |||||||| |||||||||||||| |||
Db   1459 CCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTCGATGTGATTGGGTTCCTG 1518

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||||||||||||| ||||||||||||||||||  || | ||||||||||||||||||
Db   1519 CTGGCCTGTGTGGCCACTGTGATATTTATCATCATGAAATGTTGTCTGTTTTGTTTCTGG 1578

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          |||||||||||||||| |||||||||||||| |||||||| | ||||| ||||||||||
Db   1579 AAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTGTATCTGAGATTTGAAGCT 1638

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          |||||||| |||||||  ||| ||||||||||||||||| |||     |||| || | ||
Db   1639 GGAAAACCTGATAGATGAGACTACTTCAGTTTATTCCAGGAAG-----AAAGGTTATGAT 1693

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaa 1722
          |||||||||||||||||||||||| ||||||| ||||| |||
Db   1694 GCAAGATTTCTTTCTTCCTGTGACCAAAAAAATAAAAAGAAA 1735


RESULT   7
HSAJ5162
LOCUS       HSAJ5162     2079 bp    mRNA            PRI       24-JUN-1998
DEFINITION  Homo sapiens mRNA for UDP-glucuronosyltransferase.
ACCESSION   AJ005162
VERSION     AJ005162.1  GI:3135024
KEYWORDS    UDP-glucuronosyltransferase; UGT2B4 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2079)
  AUTHORS   Ritter,J.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) Ritter J.K., Pharmacology and Toxicology,
            Virginia Commonwealth University, P.O. Box 980613, Medical College
            of Virginia, Richmond, Virginia, 23112-0613, USA
REFERENCE   2  (bases 1 to 2079)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Lubet,R.A. and Owens,I.S.
  TITLE     Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
            log differences in activity toward parallel substrates including
            hyodeoxycholic acid and certain estrogen derivatives
  JOURNAL   Biochemistry 31 (13), 3409-3414 (1992)
  MEDLINE   92207964
REFERENCE   3  (bases 1 to 2079)
  AUTHORS   Jackson,M.R., McCarthy,L.R., Harding,D., Wilson,S., Coughtrie,M.W.
            and Burchell,B.
  TITLE     Cloning of a human liver microsomal UDP-glucuronosyltransferase
            cDNA
  JOURNAL   Biochem. J. 242 (2), 581-588 (1987)
  MEDLINE   87241362
FEATURES             Location/Qualifiers
     source          1. .2079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UDPGTh-1"
                     /clone_lib="Lambda bacteriophage gt11"
                     /tissue_lib="Liver"
     gene            34. .1620
                     /gene="UGT2B4"
     CDS             34. .1620
                     /gene="UGT2B4"
                     /EC_number="2.4.1.17"
                     /function="transfers glucuronic acid to hyodeoxycholic
                     acid and estrogen catechols"
                     /codon_start=1
                     /evidence=experimental
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="CAA06396.1"
                     /db_xref="GI:3135025"
                     /db_xref="SPTREMBL:O60867"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
     conflict        546
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="g"
     conflict        548. .549
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="ct"
     conflict        1187. .1188
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="ct"
     conflict        1190
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="c"
     conflict        1192
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="a"
     conflict        1977. .1978
                     /citation=[3]
                     /replace="cga"
     conflict        2068
                     /citation=[3]
                     /replace="g"
BASE COUNT      634 a    396 c    436 g    613 t
ORIGIN


  Query Match           79.9%;  Score 1375.6;  DB 92;  Length 2079;
  Best Local Similarity  88.7%;  Pred. No. 0;
  Matches 1525;  Conservative    0;  Mismatches  189;  Indels    6;  Gaps    3;

Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
           |||| ||||||| || |||||||||||||| |||||||||||||| || || ||||||
Db      25 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTAC 84

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
           |||||||||||||||||||||||| |||||||||||| || |||||| |||||| ||||||
Db      85 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATG 144

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
           ||||| |||||||||||| | || ||||| |||||||||||||||||||||||| |||||
Db     145 AATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCA 204

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
           |||||||||||||||  ||| ||||||||      |||| ||||||||||||||||||||
Db     205 TCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTAT 264

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
           |||  |||||||||||||||| |||||| |||| |||| ||| | ||||||||||||||
Db     265 CCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGG 324
```

```
Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
Db    325 GCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGG 384

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db    385 ACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATG 444

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db    445 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGT 504

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db    505 GAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGC 564

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db    565 TACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTT 624

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db    625 ATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTG 684

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db    685 CTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGT 744

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db    745 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTT 804

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db    805 ATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTT 864

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
Db    865 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAG 924

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db    925 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCA 984

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db    985 GAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1044

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db   1045 AGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATA 1104

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db   1105 CCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1164

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
Db   1165 AATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGTTTGCA 1224

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db   1225 GATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGTTTGGACTTC 1284

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db   1285 CACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAATGATCCTTTA 1344

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db   1345 TATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTGAAGCCCCTT 1404

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
Db   1405 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1464

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db   1465 GTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACTGGGTTCCTG 1524

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
Db   1525 CTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTTTGTGTCTGG 1581

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
Db   1582 AAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGGCTGGAAGCT 1641

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
Db   1642 GGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACGTTGTGATAC 1698

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaa 1720
Db   1699 AAGAGATTCCTTTCTTCTTGTGACAAAACATCTTTCAAAA 1738
```

```
RESULT   8
AF064200
LOCUS       AF064200     2097 bp    mRNA            PRI       11-DEC-1998
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 2B4 precursor (UGT2B4)
            mRNA, UGT2B4*E458 allele, complete cds.
ACCESSION   AF064200
VERSION     AF064200.1  GI:3153831
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2097)
  AUTHORS   Levesque,E., Beaulieu,M., Hum,M. and Belanger,A.
  TITLE     Characterization and substrate specificity of UGT2B4(E458): A
            UDP-glucuronosyltransferase encoded by a polymorphic gene
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 2097)
  AUTHORS   Levesque,E., Beaulieu,M. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Ste-Foy, PQ G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2097
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .2097
                     /gene="UGT2B4"
                     /allele="UGT2B4*E458"
     CDS             26. .1612
                     /gene="UGT2B4"
                     /note="E458 variant"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B4 precursor"
                     /protein_id="AAC95002.1"
                     /db_xref="GI:3153832"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLERAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
BASE COUNT      653 a    393 c    435 g    616 t
ORIGIN
```

Query Match 79.7%; Score 1372.4; DB 88; Length 2097;
Best Local Similarity 88.5%; Pred. No. 0;
Matches 1523; Conservative 0; Mismatches 191; Indels 6; Gaps 3;

```
Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
Db      17 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTAC 76

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
Db      77 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATG 136

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
Db     137 AATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCA 196

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
Db     197 TCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTAT 256

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
Db     257 CCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGG 316

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
Db     317 GCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGG 376

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db     377 ACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATG 436

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db     437 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGT 496

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db     497 GAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGC 556

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db     557 TACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTT 616

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db     617 ATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTG 676

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db     677 CTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGT 736

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db     737 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTT 796

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db     797 ATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTT 856

Qy     841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
Db     857 CGAGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAG 916

Qy     901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db     917 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCA 976

Qy     961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db     977 GAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1036

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db    1037 AGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATA 1096

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db    1097 CCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1156

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
Db    1157 AATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGTTTGCA 1216

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db    1217 GATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGTTTGGACTTC 1276

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db    1277 CACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAATGATCCTTTA 1336

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db    1337 TATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTGAAGCCCCTT 1396

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
Db    1397 GAAAGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1456

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db    1457 GTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACTGGGTTCCTG 1516

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
Db    1517 CTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTTTGTGTCTGG 1573

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
Db    1574 AAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGGCTGGAAGCT 1633

Qy    1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
Db    1634 GGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACGTTGTGATAC 1690

Qy    1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720
Db    1691 AAGAGATTCCTTTCTTCTTGTGACAAAACATCTTTCAAAA 1730
```

RESULT 9
AF294901
LOCUS AF294901 1662 bp mRNA PRI 05-OCT-2000
DEFINITION Macaca mulatta UDP-Glucuronosyltransferase UGT2B9*2 mRNA, complete cds.
ACCESSION AF294901
VERSION AF294901.1 GI:10644704
KEYWORDS .
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 1662)
AUTHORS Dean,B., Zhao,S. and King,C.
TITLE Nucleotide and Deduced Amino Acid Sequence of UGT2B9*2: a Novel UDP-Glucuronosyltransferase Isoform Cloned from Female Rhesus Monkey Liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1662)
AUTHORS Dean,B., Zhao,S. and King,C.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2000) Drug Metabolism, Merck & Co., Mail Drop RY80L-109, PO Box 2000, Rahway, NJ 07065, USA

```
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /sex="female"
                     /tissue_type="liver"
     CDS             43. .1632
                     /note="UGT"
                     /codon_start=1
                     /product="UDP-Glucuronosyltransferase UGT2B9*2"
                     /protein_id="AAG21377.1"
                     /db_xref="GI:10644705"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENISMQEVKRWIELPK
                     DTFWLYFSQMQEIMWRFGDIIRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYIFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     SFDFYFQMYDMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMVTNMEEERANVIASALAQIPQK
                     VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTKGAAVRLDFDTMSSTDLANALKTVINDPLYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVILVI
                     MKCCLLCFWKFARKGKKGKSD"
BASE COUNT      479 a    319 c    367 g    497 t
ORIGIN
```

Query Match 79.5%; Score 1369.6; DB 88; Length 1662;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
           |||||||||||| || ||||||||||||||||  |||||| ||||| || || | ||||
Db      34 TGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACTGAGCTTTTAC 93

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
           |||||||||||||||||||||| |||||||||||||| |||||||| ||||||||||||
Db      94 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATG 153

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
           |||||||||||||||| |||||||||| ||||||||||||||||||||||||| ||||||
Db     154 AATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTCCTGGCA 213

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
           |||||||||||||||||||||||||||||  | ||||| |||||||| ||||||| | |
Db     214 TCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTCTTT 273

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
           |||||||||||||||||||| |||||||||||  ||||||| ||||||||||||||||
Db     274 CCTACATCTTTAACTAAAACTGAGTTTGAGAATATCTCCATGCAAGAGGTTAAGAGATGG 333

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
             |||  ||| |||||||| |||||||||||||||||||   |||||||| |||||
Db     334 ATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAATCATGTGG 393

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
             ||   ||||||| |||||||| |||||||||||||||||||||||||||| | |||
Db     394 AGGTTTGGTGACATAATTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATG 453

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
           || ||| |||||||||||||||||| |||||||||||||| || ||||||||||| ||
Db     454 AAGAAATTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTCCCTGTAGT 513

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
           |||||||||||| |||| |||||||||  |||||||||||||||||||| |||||||||
Db     514 GAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCACTCCTGGC 573

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
           ||||  |||||| |||  |||||||  |  |||||||||| |||||| ||||| |||||
Db     574 TACATTTTTGAAAAGCATTGTGGAGGTTTTCTTTTCCCTCCCTCCTACGTACCTGTTGTT 633

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     634 ATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATATG 693

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
           |||| ||||||||||  || ||||||| ||| |||||||||||||||| ||||| |||
Db     694 CTTTCTTTTGACTTTTATTTTCAAATGTATGACATGAAGAAGTGGGATCAATTTTATAGT 753

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
           |||||| |||||||||||||||| |||| |||||||||||| ||||||||||||||||||
Db     754 GAAGTTCTAGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACATATGGCTT 813

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
           || |||||||||||| |||||| ||||| |||||| |||||||||| ||||||||||||
Db     814 ATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTGTT 873

Qy     841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
           |||||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     874 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTTGTACAG 933

Qy     901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
           |||||||||||||||||| |||||||||||||||||||||||| || | | ||||||| |
Db     934 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCAATGGTCACTAACATGGAA 993

Qy     961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
           | |||||||||||| |||||||| |||||||||| |||||||||||||||||||||||||
Db     994 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTGTGG 1053

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
           |||||||| ||||| ||||| ||| ||||||| |||||||||||||| || |||||||||
Db    1054 AGATTTGATGGGAAAAAACCGGATACCTTAGGCCTCAATACTCGGCTCTATAAGTGGATA 1113

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    1114 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGAGCC 1173

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
           |||||||||| || || |||||||||| |||||||||||||||||||||||||||||||
Db    1174 AATGGCATCTACGAAGCGATCTACCATGGGGTCCCTATGGTGGGCATTCCATTGTTTGCC 1233

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
           ||||| |||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db    1234 GATCAGCCTGATAACATTGCTCACATGAAGACCAAAGGAGCAGCTGTTAGATTGGACTTT 1293

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
           |||||||||||||||||||| ||  |||||||||||||||||||||||||||||||||||
Db    1294 GACACAATGTCGAGTACAGACTTGGCGAATGCACTGAAGACAGTAATTAATGATCCTTTA 1353

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||||
Db    1354 TATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTG 1413

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||||
Db    1414 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGG 1473

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    1474 CCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTGGGTTCCTG 1533

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
           |||||||||||||||||||||||||||||||| || | |||| || | ||||||| ||||||||||
Db    1534 CTGGCCTGTGTGGCAACTGTGATACTTGTTATCATGAAATGTTGTCTGCTTTGTTTCTGG 1593

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1608
           |||||||||||||||| |||||||||||||| |||||||||| ||||
Db    1594 AAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTG 1641
```

```
RESULT  10
AF294902
LOCUS       AF294902     1662 bp    mRNA            PRI       05-OCT-2000
DEFINITION  Macaca mulatta UDP-Glucuronosyltransferase UGT2B33 mRNA, complete
            cds.
ACCESSION   AF294902
VERSION     AF294902.1  GI:10644706
```

```
KEYWORDS    .
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Nucleotide and Deduced Amino Acid Sequence of UGT2B33: a Novel
            UDP-Glucuronosyltransferase Isoform Cloned from Female Rhesus
            Monkey Liver
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2000) Drug Metabolism, Merck & Co., Mail Drop
            RY80L-109, PO Box 2000, Rahway, NJ 07065, USA
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /sex="female"
                     /tissue_type="liver"
     CDS             43. .1632
                     /note="UGT"
                     /codon_start=1
                     /product="UDP-Glucuronosyltransferase UGT2B33"
                     /protein_id="AAG21378.1"
                     /db_xref="GI:10644707"
                     /translation="MSVKWTSIILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENIIRQQIKRWSELPK
                     DTFWLYFSQIQEIMWRFGDISIKFCKDVVSNKKLMKKLQESRFDVVLADPIFPCSELL
                     AELFNIPLVYSLRFTPGYVFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYVL
                     YFDFCFQLYDMKKWDQFYSEVLGRHTTLSEIMGKADIWLIRNSWNFQFPHPLLPNVDF
                     IGGLLCKPAKPLPKEMEEFVQSSGENGVVVFTLGSMITNMKEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGVNTRLYKWIPQNDLLGHPKTKAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTRGAAVQLDFDTMSSTDLANALKTVINDPLYKENVMKLSRIQR
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     MKCCLFCFWKFTRKGKKGKSD"
BASE COUNT      486 a    316 c    364 g    496 t
ORIGIN
```

Query Match 79.4%; Score 1368; DB 88; Length 1662;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1458; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
Db      34 TGCACCAGGATGTCTGTGAAATGGACTTCAATAATTCTGCTAATACAACTGAGCTTTTAC 93

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
Db      94 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATG 153

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
Db     154 AATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGGGGTCATGAGGTGACTGTACTGGCA 213

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
Db     214 TCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTTTTT 273

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
Db     274 CCTACATCTTTAACTAAAACGGAGTTTGAGAATATCATCAGGCAACAGATTAAGAGATGG 333

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
Db     334 TCAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATACAAGAAATCATGTGG 393

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420

Db     394 AGGTTTGGTGACATAAGTATAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATG 453

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db     454 AAGAAACTACAAGAGTCAAGATTTGATGTCGTTTTAGCAGATCCTATTTTTCCCTGTAGT 513

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db     514 GAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCACTCCTGGC 573

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db     574 TACGTTTTTGAAAAACATTGTGGAGGATTTCTTTTCCCTCCCTCCTACGTACCTGTTGTT 633

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db     634 ATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTG 693

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db     694 CTTTATTTTGACTTTTGTTTCCAATTGTATGACATGAAGAAGTGGGATCAGTTTTATAGT 753

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db     754 GAAGTTCTAGGAAGACACACTACATTATCTGAGATAATGGGGAAAGCTGACATATGGCTT 813

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db     814 ATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTATT 873

Qy     841 ggaggattccactgcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 900
Db     874 GGAGGACTCCTCTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAATTTGTACAG 933

Qy     901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db     934 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTACTCTGGGGTCAATGATCACTAACATGAAA 993

Qy     961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db     994 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATTCCACAAAAGGTTCTGTGG 1053

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db    1054 AGATTTGATGGGAATAAACCAGATACCTTAGGTGTCAATACTCGGCTCTATAAGTGGATA 1113

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db    1114 CCCCAGAATGACCTTCTAGGTCATCCAAAAACCAAAGCTTTTATAACTCACGGTGGAGCC 1173

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
Db    1174 AATGGCATCTACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCACTGTTTGCC 1233

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db    1234 GATCAACCTGATAACATTGCTCACATGAAAACCAGAGGAGCAGCTGTTCAATTGGACTTC 1293

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db    1294 GACACAATGTCGAGTACAGACTTGGCGAATGCACTGAAGACAGTAATTAATGATCCTTTA 1353

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db    1354 TATAAAGAGAATGTTATGAAATTATCAAGAATTCAACGTGATCAGCCAGTGAAGCCCCTG 1413

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
Db    1414 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGG 1473

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db    1474 CCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTGGGTTCCTG 1533
```

```
Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||||||||||||||| ||||||||||||||||||||  || | ||||||||||||||||||
Db   1534 CTGGCCTGTGTGGCCACTGTGATATTTATCATCATGAAATGTTGTCTGTTTTGTTTCTGG 1593

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1608
          |||||| |||||||||| |||||||||||||| |||||||||| ||||
Db   1594 AAGTTTACTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTG 1641


RESULT 11
AF081793
LOCUS       AF081793     2091 bp    mRNA            PRI       18-AUG-1998
DEFINITION  Homo sapiens UDP glucuronosyltransferase 2B4 precursor (UGT2B4)
            mRNA, UGT2B4*L109, L396 variant allele, complete cds.
ACCESSION   AF081793
VERSION     AF081793.1  GI:3426331
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2091)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
  MEDLINE   93326164
REFERENCE   2  (bases 1 to 2091)
  AUTHORS   Mackenzie,P.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1998) Clinical Pharmacology, Flinders University
            of South Australia, Bedford Park, SA 5042, Australia
FEATURES             Location/Qualifiers
     source          1. .2091
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
     gene            1. .2091
                     /gene="UGT2B4"
                     /note="formerly UGT2B11"
                     /allele="UGT2B4*L109, L396 variant"
     CDS             36. .1622
                     /gene="UGT2B4"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 2B4 precursor"
                     /protein_id="AAC32272.1"
                     /db_xref="GI:3426332"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYLSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKKWKEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLLADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
BASE COUNT      638 a    401 c    437 g    615 t
ORIGIN


  Query Match             79.3%;  Score 1366;  DB 88;  Length 2091;
  Best Local Similarity   88.3%;  Pred. No. 0;
  Matches 1519;  Conservative    0;  Mismatches  195;  Indels    6;  Gaps    3;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||| |||||| || |||||||||||||| |||||||||||||| || || ||||||
Db     27 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTAC 86

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          |||||||||||||||||||| |||||||||||| || |||||| |||||| ||||||
Db     87 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATG 146

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          ||||| |||||||||||| | || ||||| |||||||||||||||||||||||||| |||||
Db    147 AATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCA 206

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          |||||||||||||||  ||| ||||||||     |||| ||||||||||||||||||||
Db    207 TCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTAT 266

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          |||  |||||||||||||||||| |||||| |||| |||| ||| | ||||||||||||
Db    267 CCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGG 326

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          ||||  ||| |||||| |  ||||||| |||| ||||||||| |||||||||| |||||
Db    327 GCAGAACTTCCAAAAGACACATTTTGGTCATATCTTTCACAAGTACAAGAAATCATGTGG 386

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
            |||  |||||||| ||||||| |||||||| ||| ||||||||||||||||| |||||
Db    387 ACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATG 446

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          || |||||||| |||||||||||||  | ||| ||||||| |||||||| |||| ||||
Db    447 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGACGCTGTTTTCCCCTTTGGT 506

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||  | | ||||||| ||||  ||||| ||||| ||||||||  ||||||||
Db    507 GAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGC 566

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||| |||||||||  ||| ||||||||||  | ||||||||||||||  | ||| ||||||
Db    567 TACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTT 626

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| || ||||||| |||||||||||||| ||||||||||||||||||||||||||||
Db    627 ATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTG 686

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          |||||||||| |||||||||||||| | ||| ||||||||||||||||||||| ||||||
Db    687 CTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGT 746

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||||| |||| |||||||||| ||||||||||||||||||||
Db    747 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTT 806

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || ||||||| |||| |||||||||||||| ||| |||||||||| ||||| || |||
Db    807 ATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTT 866

Qy    841 ggaggattccactgcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||||| ||| ||  | |||||||||| |||
Db    867 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGAAATGGAAGGAGTTTGTCCAG 926

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          ||||||||||||||||||||||||||||||||||||||||||||||  || | ||||||| | ||
Db    927 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCA 986

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    987 GAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1046

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||| | ||||| ||||||||||||||||||||||||||||
Db   1047 AGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATA 1106

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||| |||| ||||| ||||||||||||||||||||||||||||||||||||||
Db   1107 CCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1166

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1167 AATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGCTTGCA 1226
```

```
Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          ||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
Db   1227 GATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGTTTGGACTTC 1286

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          |||||||||||||||||||| | || ||||||||||||||||||||||||||||||||||
Db   1287 CACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAATGATCCTTTA 1346

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          |||||||||||  |||||||||||||||||||  ||||||||||||| ||||||||
Db   1347 TATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTGAAGCCCCTT 1406

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          ||||||||||||||||||||||||||||||||||| ||| |||||||||||| ||||||||
Db   1407 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1466

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
          |||||||||||  |||||||||||||||||||||||||||||||||||||||  |||||| |||
Db   1467 GTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACTGGGTTCCTG 1526

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||||||||||||||||||||||||  ||||||||||    |||||||||||  |||||
Db   1527 CTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTTTGTGTCTGG 1583

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          |||||||  |||||  ||| |||||||| ||||||||||  ||| ||||||    | ||||||
Db   1584 AAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGGCTGGAAGCT 1643

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          ||  |||||  |||  ||  |  ||  ||| |||||||| ||  ||||||    |      ||  ||
Db   1644 GGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACGTTGTGATAC 1700

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720
             |||||  ||||||||  ||||||||| |       ||||
Db   1701 AAGAGATTCCTTTCTTCTTGTGACAAAACATCTTTCAAAA 1740
```

```
RESULT  12
HSUDPGT
LOCUS       HSUDPGT      2093 bp    mRNA            PRI       23-MAR-1995
DEFINITION  Human mRNA for liver microsomal UDP-glucuronosyltransferase
            (UDPGT).
ACCESSION   Y00317
VERSION     Y00317.1  GI:37588
KEYWORDS    transferase; UDP-glucuronosyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2093)
  AUTHORS   Burchell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1987) Burchell B., Dept. Biochemistry, Medical
            Sciences Institute, University Dundee, Scotland
REFERENCE   2  (bases 1 to 2093)
  AUTHORS   Jackson,M.R., McCarthy,L.R., Harding,D., Wilson,S., Coughtrie,M.W.
            and Burchell,B.
  TITLE     Cloning of a human liver microsomal UDP-glucuronosyltransferase
            cDNA
  JOURNAL   Biochem. J. 242 (2), 581-588 (1987)
  MEDLINE   87241362
FEATURES             Location/Qualifiers
     source          1. .2093
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11 liver cDNA"
                     /clone="Hlug25"
     sig_peptide     38. .106
                     /product="put. signal peptide"
     CDS             38. .1624
                     /note="precursor"
                     /codon_start=1
                     /protein_id="CAA68415.1"
                     /db_xref="GI:37589"
                     /db_xref="SWISS-PROT:P06133"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYRPRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYKAISPRIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
     mat_peptide     107. .1621
                     /product="mature UDPGT (AA 1-505) (EC 2.4.1.17)"
     misc_feature    979. .988
                     /note="pot. glycosylation site"
     misc_feature    2076. .2081
                     /note="pot. polyA signal"
BASE COUNT      638 a    397 c    439 g    619 t
ORIGIN
```

Query Match 79.2%; Score 1364.4; DB 93; Length 2093;
Best Local Similarity 88.3%; Pred. No. 0;
Matches 1518; Conservative 0; Mismatches 196; Indels 6; Gaps 3;

```
Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||| ||||||| || |||||||||||||||| |||||||||||||||| || ||||||
Db     29 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTAC 88

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          |||||||||||||||||||||| |||||||||| || |||||| |||||| ||||||
Db     89 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATG 148

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          ||||| |||||||||||| | || ||||| |||||||||||||||||||||||| |||||
Db    149 AATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCA 208

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          ||||||||||||||  ||| ||||||||    |||| ||||||||||||||||||||||
Db    209 TCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTAT 268

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          |||  |||||||||||||||||| |||||| |||| |||| ||| | ||||||||||||
Db    269 CCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGG 328

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          ||||  ||| |||||| |  ||||||| |||||||||||||| |||||||||| |||||
Db    329 GCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGG 388

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
          |||  |||||||| ||||||| |||||||| ||| ||||||||||||||| ||||||
Db    389 ACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATG 448

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          || |||||||| ||||||||||||||  | ||| ||||||||||||||| |||| ||||
Db    449 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGT 508

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||  | | ||||||| ||||  ||||| ||||| |  |||||  ||||||||
Db    509 GAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGGCCTCGCTTCTCTCCTGGC 568

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||| |||||||||  ||| ||||||||||  | ||||||||||||||  | ||| |||||
Db    569 TACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTT 628

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| || ||||||| |||||||||||||| ||||||||||||||||||||||||||
Db    629 ATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTG 688

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          |||||||||||  |||||||||||||| | ||| ||||||||||||||||||||| ||||||
```

```
Db    689 CTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGT 748

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||||| |||| |||||||||| |||||||||||||||||||
Db    749 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTT 808

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || ||||||| |||| ||||||||||||||| ||| |||||||||| ||||| || |||
Db    809 ATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTT 868

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||||| |||||||||| |||||||| |||
Db    869 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAG 928

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          |||||||||||||||||||||||||||||||||||||||||| || | ||||||| | ||
Db    929 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCA 988

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    989 GAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1048

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||| | ||||| ||||||||||||||||||||||||||||
Db   1049 AGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATA 1108

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||||
Db   1109 CCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCC 1168

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
          |||||||||||| |||||||||  | | | ||||||||||||||| ||||||||||
Db   1169 AATGGCATCTATAAGGCAATCTCTCCTAGAATCCCTATGGTGGGCGTTCCATTGTTTGCA 1228

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          |||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
Db   1229 GATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGTTTGGACTTC 1288

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          |||||||||||||||||||| | || ||||||||||||||||||||||||||||||||||
Db   1289 CACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAATGATCCTTTA 1348

Qy   1321 tataaagagaatattatgaaatttatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          |||||||||||  |||||||||||||||||||| |||||||||||||| ||||||||
Db   1349 TATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTGAAGCCCCTT 1408

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          |||||||||||||||||||||||||||||||||||| ||| ||||||||||| ||||||||
Db   1409 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1468

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
          |||||||||||| ||||||||||||||||||||||||||||||||||||||| ||| |||
Db   1469 GTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACTGGGTTCCTG 1528

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||||||||||||||||||||||||| ||||||||||    |||||||||||| |||||
Db   1529 CTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTTTGTGTCTGG 1585

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          ||||||| ||||| |||| |||||||| |||||||||| ||| ||||||   | ||||||
Db   1586 AAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGGCTGGAAGCT 1645

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          || |||||  ||| || | ||  ||| |||||||| || ||||||   |     || ||
Db   1646 GGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACGTTGTGATAC 1702

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720
             ||||| |||||||| |||||||||| |      ||||
Db   1703 AAGAGATTCCTTTCTTCTTGTGACAAAACATCTTTCAAAA 1742


RESULT 13
AF112112
```

```
LOCUS       AF112112     2108 bp    mRNA            PRI       11-APR-1999
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase 2B19 precursor,
            mRNA, complete cds.
ACCESSION   AF112112
VERSION     AF112112.1  GI:4580601
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryotä; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Molecular cloning, expression and characterization of a monkey
            steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates
            testosterone
  JOURNAL   Eur. J. Biochem. 260 (3), 701-708 (1999)
  MEDLINE   99203465
REFERENCE   2  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2108
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             26. .1612
                     /EC_number="2.4.1.17"
                     /note="UGT2B19"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B19 precursor"
                     /protein_id="AAD24435.1"
                     /db_xref="GI:4580602"
                     /translation="MSMKWTSALLLIQLSCYLSFGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLAYSTSILPDPNNPSPLKFEICPTSLTETEFQDSVTQLVKRWSDIRK
                     DTFWPHFLHVQEMMWTYGDMIRKFCKDVVSNKKLMKKLQESRFDVVLADAISPCGELL
                     AELLKIPFVYSLRFSPGYALEKHGGGFLFPPSYVPVTMSELRDQMTFMERVQNMIYMV
                     YFDFWFQVWDVKNWDQFYSKVLGRPTTLFEIMAKAEIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGDNGVVVFSLGSMVSNMSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTQLYKWLPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVRLDFDTMSSTDLLNALKTVINDPIYKENAMKLSSIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     TKCLFCVWKFVRTRKKGKRD"
BASE COUNT      649 a    393 c    438 g    628 t
ORIGIN


  Query Match           78.5%;  Score 1351.2;  DB 88;  Length 2108;
  Best Local Similarity  87.4%;  Pred. No. 2.5e-296;
  Matches 1504;  Conservative    0;  Mismatches  208;  Indels    8;  Gaps    2;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||| ||||||| || |||||||||||||| |||||||||||||| || || ||||||
Db     17 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTAC 76

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          || |||| |||||||||||||| |||||||||||| || |||||| ||||||||||||||
Db     77 TTGAGCTTTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCATTGGATG 136

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          ||||| |||||||| ||| | ||||||| |||||||||||||||||||||||||||||||
Db    137 AATATAAAGACAATTCTGGATGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTACTGGCA 196

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          | |||| |||||||||||  |||||||||  |  ||||  |||||||||||||| ||| |
Db    197 TATTCAACTTCCATTCTTCCTGATCCCAACAACCCATCTCCTCTTAAATTTGAAATTTGT 256

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          |||||||||||||  ||||||| ||| || ||| | ||| ||||| ||||||||||||||
Db    257 CCTACATCTTTAACCGAAACTGAGTTTCAGGATAGCGTCACGCAACTGGTTAAGAGATGG 316

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
```

```
Db    317 TCAGACATTCGAAAAGACACATTTTGGCCACATTTTTTACACGTACAAGAAATGATGTGG 376

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db    377 ACATATCGTGACATGATTAGAAAGTTCTGTAAAGATGTGGTTTCTAATAAGAAACTTATG 436

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db    437 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCGGATGCTATTTCCCCCTGTGGT 496

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db    497 GAGCTACTGGCTGAGCTACTTAAAATACCCTTTGTGTACAGCCTCCGCTTCTCTCCTGGC 556

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db    557 TACGCACTTGAAAAGCATGGTGGAGGGTTTCTGTTCCCTCCTTCCTATGTACCTGTTACT 616

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db    617 ATGTCAGAATTAAGAGATCAAATGACTTTCATGGAGAGAGTACAAAATATGATCTATATG 676

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db    677 GTTTATTTTGACTTTTGGTTCCAAGTATGGGATGTAAAGAACTGGGATCAGTTTTACAGT 736

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db    737 AAAGTTCTAGGAAGACCCACTACATTATTTGAGATAATGGCAAAAGCTGAAATATGGCTT 796

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db    797 ATTCGAAACTACTGGGATTTTCAATTTCCTCATCCACTCTTACCAAATGTTGAGTTCGTT 856

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
Db    857 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTCGTCCAG 916

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db    917 AGCTCTGGAGACAATGGTGTTGTGGTGTTTTCTCTGGGGTCCATGGTCAGTAACATGTCA 976

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db    977 GAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1036

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db   1037 AGATTTGATGGGAATAAACCAGATACCTTAGGACTCAATACTCAGCTGTACAAGTGGCTA 1096

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db   1097 CCCCAGAATGATCTTCTTGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGAGCC 1156

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
Db   1157 AATGGCATCTACGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGTTTGCC 1216

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db   1217 GATCAACCTGATAACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTTAGACTGGACTTC 1276

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db   1277 GACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATAAATGATCCTATA 1336

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db   1337 TATAAAGAGAATGCTATGAAATTATCAAGCATTCATCATGATCAACCAGTGAAGCCCCTG 1396

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
```

```
Db   1397 GATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAACACCTTCGG 1456

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db   1457 GTTGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTG 1516

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
Db   1517 CTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTTTGTGTCTGG 1573

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
Db   1574 AAGTTTGTTAGAACAAGAAAGAAGGGAAAAAGAGATTAAATACATCTGAGGCTGGAAGCT 1633

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
Db   1634 GGGAAACCTGATAAATGAGATTTCTTCAGTTCATAC-----AATGAGAATAGGTTGTGAT 1688

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720
Db   1689 ACAAGATTCCTTTCTTCTTGTGCCAAAACATCTTTCAAAA 1728
```

```
RESULT  14
AX040087
LOCUS       AX040087     1650 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 10 from Patent WO0063351.
ACCESSION   AX040087
VERSION     AX040087.1  GI:11230049
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Baughn,M.R. and Yang,J.
  TITLE     Carbohydrate-modifying enzymes
  JOURNAL   Patent: WO 0063351-A 10 26-OCT-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2912330CB1"
BASE COUNT      489 a    330 c    354 g    477 t
ORIGIN
```

```
  Query Match           71.9%;  Score 1238;  DB 9;  Length 1650;
  Best Local Similarity  85.7%;  Pred. No. 1.2e-270;
  Matches 1376;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
Db     26 TGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTCCTGATACAGCTGAGCTGTTAC 85

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
Db     86 TTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATG 145

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
Db    146 AATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCA 205

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
Db    206 TCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTAT 265

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
Db    266 CCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGG 325

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
```

```
Db    326 GCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGG 385

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
          |||  |||||||| ||||||| |||||||| ||| |||||||||||||||| |||||
Db    386 ACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATG 445

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          || |||||||| |||||||||||||  | ||| |||||||||||||||| |||| ||||
Db    446 AAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGT 505

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||| | | ||||||| |||| ||||| ||||| ||||||||  ||||||||
Db    506 GAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGC 565

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||| |||||||| ||| |||||||||  | ||||||||||||||  | ||| |||||
Db    566 TACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTT 625

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| || ||||||| |||||||||||||| ||||||||||||||||||||||||||
Db    626 ATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTG 685

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          |||||||||| |||||||||||||| | ||| ||||||||||||||||||||| ||||||
Db    686 CTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGT 745

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||| |||| |||||||||| ||||||||||||||||||||
Db    746 GAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTT 805

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || ||||||| |||| |||||||||||||| ||| |||||||||| ||||| || |||
Db    806 ATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTT 865

Qy    841 ggaggattccactgcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||||||||| |||||||||| |||||||| |||
Db    866 GGAGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAG 925

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          ||||||||||||||||||||||||||||||||||||||||||  || | ||||||| | ||
Db    926 AGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCA 985

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    986 GAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGG 1045

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
           |||||||| |||||||||||||| | ||||| ||||||||||||||||||||||||||||
Db    1046 AGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATA 1105

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggaagcc 1140
           |||||||||| ||||| |||||||| ||||||| |||||||| ||||||||||
Db    1106 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1165

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
           ||||| |||||||| || || ||||||||| |||||||||||||  ||||  | |||  |
Db    1166 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1225

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
           ||||| | |||||||||  |||||||||||||||||  ||||||||||  | |  |||||
Db    1226 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1285

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctttta 1320
           || || ||| | ||   |||  | ||||  ||  |||  |||||  ||||  ||| | |
Db    1286 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1345

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
           |||||||||||  ||||| |||||||||||||| |||||||||| ||||||||||
Db    1346 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1405

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
           |||||||||||||||||||| || |||||||||| |||||||||||||||| ||||| |||

Db    1406 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1465

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           || |||||||||||||||||||||||||| || ||||| | |||||||||||||| |||
Db    1466 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1525

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
           ||| |||||||||||||| |||||  |  ||||||| | || | | ||||  |
Db    1526 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1585

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtc 1606
           || |||  || ||   | ||||  | ||| || || ||| | | ||
Db    1586 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1631


RESULT 15
HSU59209
LOCUS       HSU59209     2107 bp    mRNA            PRI       02-JUL-1998
DEFINITION  Homo sapiens C19steroid specific UDP-glucuronosyltransferase mRNA,
            complete cds.
ACCESSION   U59209
VERSION     U59209.1  GI:3287472
KEYWORDS    UGT2B17G.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2107)
  AUTHORS   Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
  TITLE     Isolation and characterization of a novel cDNA encoding a human
            UDP-glucuronosyltransferase active on C19 steroids
  JOURNAL   J. Biol. Chem. 271 (37), 22855-22862 (1996)
  MEDLINE   96394358
REFERENCE   2  (bases 1 to 2107)
  AUTHORS   Hum,D.W., Belanger,A., Beaulieu,M. and Levesque,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1996) Laboratory of Molecular Endocrinology,
            Centre Hospitalier de l'universite Laval, 2705 Boul. Laurier,
            Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="prostate"
                     /cell_type="LNCaP"
     CDS             52. .1644
                     /note="UGT2B17"
                     /codon_start=1
                     /product="C19steroid specific UDP-glucuronosyltransferase"
                     /protein_id="AAC25491.1"
                     /db_xref="GI:3287473"
                     /translation="MSLKWMSVFLLMQLSCYFSSGSCGKVLVWPTEYSHWINMKTILE
                     ELVQRGHEVIVLTSSASILVNASKSSAIKLEVYPTSLTKNDLEDFFMKMFDRWTYSIS
                     KNTFWSYFSQLQELCWEYSDYNIKLCEDAVLNKKLMRKLQESKFDVLLADAVNPCGEL
                     LAELLNIPFLYSLRFSVGYTVEKNGGGFLFPPSYVPVVMSELSDQMIFMERIKNMIYM
                     LYFDFWFQAYDLKKWDQFYSEVLGRPTTLFETMGKAEMWLIRTYWDFEFPRPFLPNVD
                     FVGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMISNMSEESANMIASALAQIPQ
                     KVLWRFDGKKPNTLGSNTRLYKWLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMV
                     GIPLFADQHDNIAHMKAKGAALSVDIRTMSSRDLLNALKSVINDPIYKENIMKLSRIH
                     HDQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWIQYHSLDVIAFLLACVATMIFM
                     ITKCCLFCFRKLAKTGKKKRD"
BASE COUNT      657 a    382 c    424 g    644 t
ORIGIN


  Query Match           70.8%;  Score 1219.2;  DB 93;  Length 2107;
  Best Local Similarity  83.7%;  Pred. No. 2.2e-266;
  Matches 1430;  Conservative    0;  Mismatches  268;  Indels   10;  Gaps    4;

Qy      4 accaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttacttt 63
          |||||||| |||||||||||  |||||  |||||||||| || ||||||||||||||
Db     46 ACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCTCAGTTGTTACTTT 105
```

```
Qy      64 agctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaat 123
           ||||||||||||||||||| |||||||||||| || ||||||||||||||||||| |||
Db     106 AGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCCATTGGATAAAT 165

Qy     124 atgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatct 183
           ||||||||||||||| |||||||||||||| |||||||||||| ||| || |||||
Db     166 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTGTGTTGACATCT 225

Qy     184 tcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcct 243
           || ||||| |||||| |  || ||| | |  ||||  || |||||||  ||||||||||||
Db     226 TCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 285

Qy     244 acatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatggtca 303
           ||||||||||||||| ||| || ||  || |  | ||| ||  | |  | ||||||  ||
Db     286 ACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCGATAGATGGACA 345

Qy     304 ga---cattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
           |    |||  |||| ||| |||||||| ||||||||||||  |||||||  |    |||
Db     346 TATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAGAATTGTGTTGG 405

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
           ||||  | ||||      || |||  ||||| |||||| |||||  || |||||| |||||
Db     406 GAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACAAGAAACTTATG 465

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
           | |||||||||||||||| ||||||  || || | |||||||| |||  ||||||||||
Db     466 AGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTAATCCCTGTGGT 525

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
           ||||||||||||  ||||||||||||  ||| ||||||||||||||||  ||  ||||
Db     526 GAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCTTCTCTGTTGGC 585

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
           |||||| |||| | | |  |||||||| |  | ||||||||||||||  |||||  |||||
Db     586 TACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 645

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
           |||||| |||||||||||||||||| ||||||||||||  ||||||||||||  ||| ||
Db     646 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATATATATG 705

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
           ||||||||||||||||||| |||   | |||| |||||||||||| ||||||||  |||
Db     706 CTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACCAGTTTTATAGT 765

Qy     721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
           ||||||  |||||||||||||||| |||||||||||||||||  |||||||| || ||||||
Db     766 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 825

Qy     781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
           || |||| ||  |||  |||| |||||||||  |||||||||||||| ||||||||||||
Db     826 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 885

Qy     841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
           ||||||  | |||||  ||||| |||||||||  | ||||||||||||||  |||||||| |||
Db     886 GGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAAGAGTTTGTGCAG 945

Qy     901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
           ||||||||||||||||||  ||||||||||||||||||||||  |||| |||||||| ||
Db     946 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 1005

Qy     961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
           | |||||| |||||  | |||||| |||||||||||  ||||||||||||||||||| |||
Db    1006 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1065

Qy    1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
           ||||||||  || || ||  ||| || | ||||||  |||||||| |||||  |||||| ||
Db    1066 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTATAAGTGGTTA 1125

Qy    1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
           |||||||||||||||||  ||||||||  |||||||  ||||||||||||||||||||||  ||
Db    1126 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1185

Qy    1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
           |||||||||||||||||  |||||||||||||||||||||||||||||||||  ||||||
Db    1186 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1245

Qy    1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
           |||||||  |||||||||||||||||||||||||||||||||  | ||  ||||| ||
Db    1246 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1305

Qy    1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
           |  || ||||| |||| |||  |||| |||||| ||||| |||| |||||||| ||| |
Db    1306 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTATC 1365

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
           |||||||||||||| ||||||||||||||||||| |||||||||| || |||||||||
Db    1366 TATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCGGTGAAGCCCCTG 1425

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
           ||||||||||||||||||||||||| |||||||||| ||| ||||||||||  ||||||||
Db    1426 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1485

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           || |||||||| |||||||||| |||||||||||||||||||||||| |  || |||
Db    1486 GTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1545

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
           |||||||| ||||||||| |||||||||| |||||||| | ||| ||||||||||| |
Db    1546 CTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTGTTTTGTTTCCGA 1605

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
           ||| |||| | || ||| ||||||   ||||| ||||||||||  ||  |   ||||| |
Db    1606 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAAAAGCCTGAAG-T 1664

Qy    1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
           ||||   |  | |||| ||||  || |  |||||||||||| | | |    | ||   ||
Db    1665 GGAATGACCAAAAGATGGGACTCCTCC--TTTATTCCAGCATGGAGG----GTTTTAAAT 1718

Qy    1681 gcaagatttctttcttcctgtgacaaaa 1708
           | | |||||| || |||||| |||||||
Db    1719 GGAGGATTTCCTTTTTCCTGCGACAAAA 1746
```

Search completed: April 30, 2001, 19:15:44
Job time: 37418 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:07:29 ; Search time 933.78 Seconds
(without alignments)
1076.561 Million cell updates/sec

Title: US-09-721-183-18
Perfect score: 1722
Sequence: 1 tgcaccaggatgactctgaa.....acaaaaaaaaaaaaaaaaaa 1722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

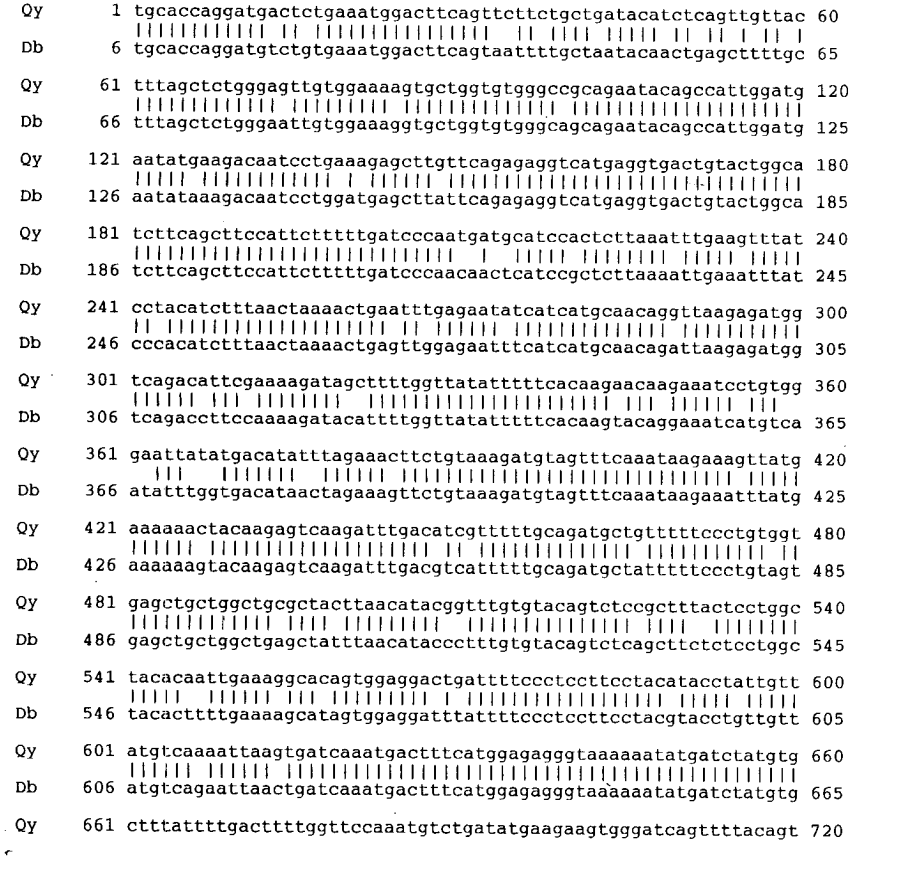
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1478	85.8	1854	21	Z95200	Human UDP-glucuron
2	1375.6	79.9	2092	21	Z95199	Human UDP-glucuron
3	1238	71.9	1650	21	C65396	Human carbohydrate
4	1219.2	70.8	2107	19	V15900	Uridine diphospho-
5	1213.6	70.5	1976	21	Z95206	Human UDP-glucuron
6	608.2	35.3	1686	21	Z95201	Human UGT2B7 exon
7	584.8	34.0	1323	21	Z95193	Human UGT2B4 exon
8	498.8	29.0	2312	21	Z95207	Human UGT2B15 exon
9	343	19.9	596	21	Z95205	Human UGT2B7 exon
10	311.4	18.1	2368	13	Q27370	HUG-Br2. Homo sap
11	302.6	17.6	2351	13	Q27369	HUG-Br1. Homo sap
12	271	15.7	1589	21	Z95198	Human UGT2B4 exon
13	262.6	15.2	735	21	Z45118	UDP-glucuronosyltr
14	250.4	14.5	978	21	Z95211	Human UGT2B15 exon
15	230	13.4	350	21	C03286	Human secreted pro
16	207.6	12.1	1591	21	Z95204	Human UGT2B7 exon
17	205.8	12.0	689	21	Z95197	Human UGT2B4 exon
18	180	10.5	1602	21	Z95210	Human UGT2B15 exon
19	177	10.3	264	21	F15707	Human prostate can
20	156	9.1	223	21	A87255	Rat hepatocyte car
21	133.4	7.7	1340	21	Z95202	Human UGT2B7 exon
22	131.6	7.6	746	21	Z95194	Human UGT2B4 exon
23	129	7.5	2320	21	C58107	Human PRO1780 nucl
24	129	7.5	2320	21	A37101	Human PRO1780 (UNQ
25	123	7.1	391	20	Z07208	Human lung tumour
26	123	7.1	391	21	C79066	Human lung tumour-
27	122	7.1	1020	21	Z95208	Human UGT2B15 exon
28	121	7.0	1822	21	Z95203	Human UGT2B7 exon
29	120	7.0	273	21	A87491	Rat hepatocyte car
30	120	7.0	785	21	Z95195	Human UGT2B4 exon
31	114.6	6.7	480	21	Z95209	Human UGT2B15 exon
32	108	6.3	283	21	A87467	Rat hepatocyte car
33	106.6	6.2	1667	13	Q33027	Exon 5 from the UG
34	103.2	6.0	515	20	V87412	EST clone BR77. H
35	100.2	5.8	164	21	A87154	Rat hepatocyte car
36	96.4	5.6	508	21	C98478	Human colon cancer
37	92.6	5.4	259	21	A87473	Rat hepatocyte car
38	84.2	4.9	1138	21	Z95196	Human UGT2B4 exon
39	81.6	4.7	775	13	Q33026	Exons 3-4 from the
40	73.2	4.3	219	21	C13974	Human secreted pro
41	73.2	4.3	366	21	A87085	Rat hepatocyte car
42	67.4	3.9	1048	13	Q33020	UGT1E Exon 1 from
43	64.8	3.8	867	21	Z45111	UDP-glucuronosyltr
44	64.8	3.8	867	21	Z45112	UDP-glucuronosyltr
45	64.8	3.8	867	21	Z45113	UDP-glucuronosyltr

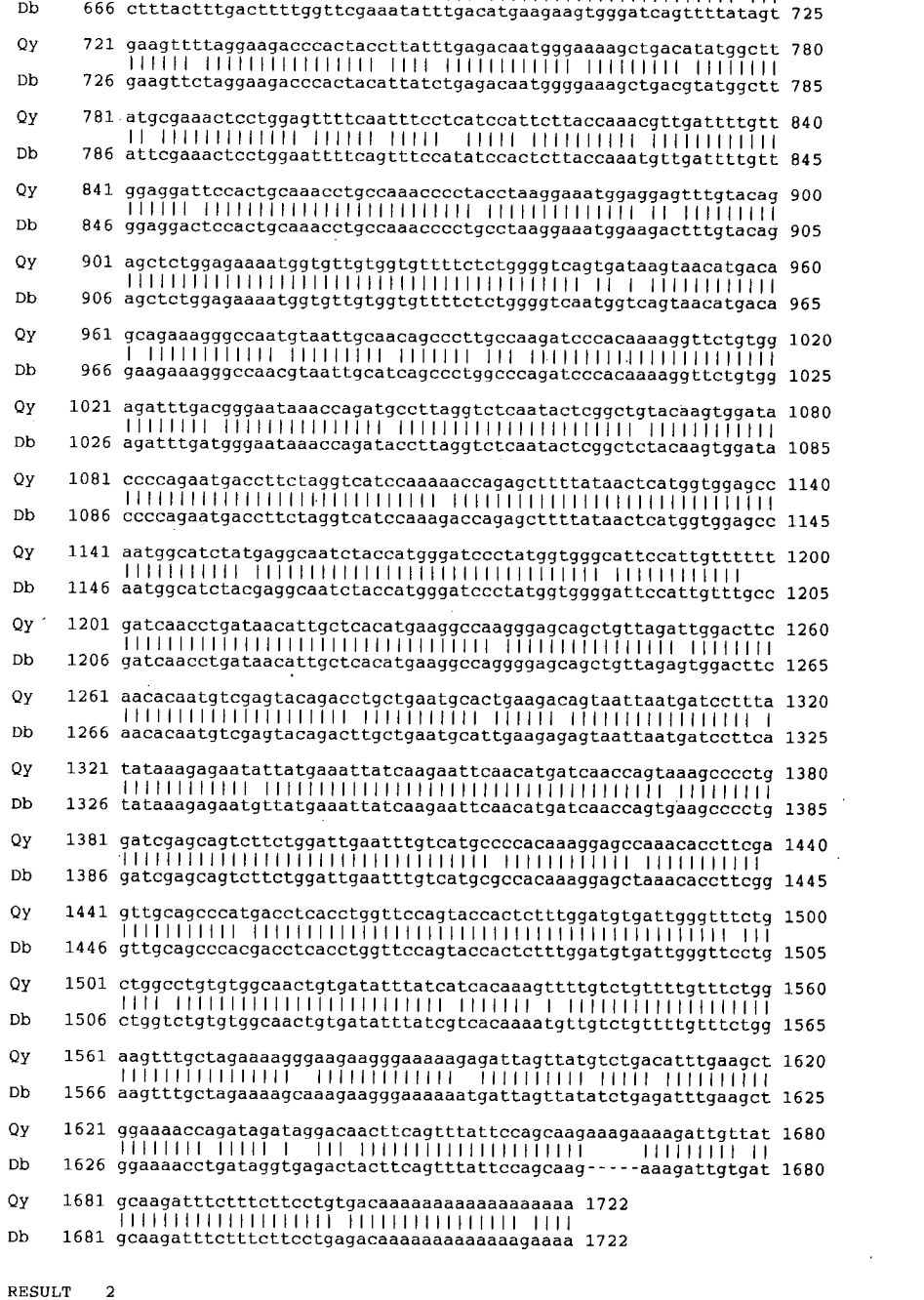
ALIGNMENTS

RESULT 1
Z95200
ID Z95200 standard; DNA; 1854 BP.
XX
AC Z95200;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B7 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; Y78934.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX

PS Disclosure; Page 41-44; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B7
CC (UGT2B7) gene. UDP-glucuronosyltransferase (UGTs) are a family of enzymes
CC that catalyse the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC substrate in an individual. The nucleic acid molecules comprising a human
CC UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of UGT2B
CC substrate, potential drug-drug interactions and adverse side effects. The
CC polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC detecting genetic linkage related to phenotypic variation in activity or
CC expression of UGT2B protein. The polymorphism containing nucleic acid
CC molecules may also be used for generating genetically modified non-human
CC animals and for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 1854 BP; 572 A; 338 C; 392 G; 552 T; 0 other;

Query Match 85.8%; Score 1478; DB 21; Length 1854;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 1577; Conservative 0; Mismatches 140; Indels 5; Gaps 1;





RESULT 2

```
Z95199
ID   Z95199 standard; DNA; 2092 BP.
XX
AC   Z95199;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B4 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   P-PSDB; Y78933.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions  -
XX
PS   Disclosure; Page 34-36; 72pp; English.
XX
CC   This sequence represents the human UDP-glucuronosyltransferase 2B4
CC   (UGT2B4) gene. UDP-glucuronosyltransferase (UGTs) are a family of enzymes
CC   that catalyse the glucuronic acid conjugation of a wide range of
CC   endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC   steroid metabolizing isoforms in the liver. Alteration of the expression
CC   or function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC   locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC   substrate in an individual. The nucleic acid molecules comprising a human
CC   UGT2B sequence polymorphism can be used in screening assays for
CC   genotyping individuals, also to predict their rate of metabolism of UGT2B
CC   substrate, potential drug-drug interactions and adverse side effects. The
CC   polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC   detecting genetic linkage related to phenotypic variation in activity or
CC   expression of UGT2B protein. The polymorphism containing nucleic acid
CC   molecules may also be used for generating genetically modified non-human
CC   animals and for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 2092 BP; 639 A; 398 C; 438 G; 617 T; 0 other;

  Query Match             79.9%;  Score 1375.6;  DB 21;  Length 2092;
  Best Local Similarity   88.7%;  Pred. No. 0;
  Matches 1525;  Conservative    0;  Mismatches  189;  Indels    6;  Gaps    3;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          |||| |||||||  || ||||||||||||||| ||||||||||||||| || ||  ||||||
Db     29 tgcatcaggatgtctatgaaatggacttcagctcttctgctgatacagctgagctgttac 88

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          |||||||||||||||||||||||| |||||||||||| || ||||||  ||||||  ||||||
Db     89 tttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagccactggatg 148

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          |||||  |||||||||||| | || ||||| ||||||||||||||||||||||||||||||
Db    149 aatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactgtattggca 208

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          |||||||||||||||  ||| ||||||||     |||| ||||||||||||||||||||||
Db    209 tcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttgaagtttat 268

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          |||  ||||||||||||||||||| |||||| |||| |||| ||| | ||||||||||||
Db    269 cctgtatctttaactaaaactgagtttgaggatattatcaagcagctggttaagagatgg 328

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          ||||  ||| |||||| | |||||||| |||||||||||||| |||||||||| |||||
Db    329 gcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaatcatgtgg 388

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
            |||  |||||||| ||||||| |||||||| ||| ||||||||||||||| |||||
Db    389 acatttaatgacatacttagaaagttctgtaaggatatagtttcaaataagaaacttatg 448

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          || |||||||| |||||||||||||  | ||| ||||||||||||||| |||| ||||
Db    449 aagaaactacaggagtcaagatttgatgttgttcttgcagatgctgttttccccttggt 508

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||  | | ||||||| ||||  ||||| ||||| ||||||||  ||||||||
Db    509 gagctgctggccgagttacttaaaataccctttgtctacagcctccgcttctctcctggc 568

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||| |||||||| ||| ||||||||||  | ||||||||||||  | ||| |||||
Db    569 tacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgcctgttgtt 628

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| || ||||||| |||||||||||||| ||||||||||||||||||||||||||||
Db    629 atgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatgatctatgtg 688

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          |||||||||| |||||||||||||| | ||| ||||||||||||||||||||| ||||||
Db    689 ctttattttgaattttggttccaaatatttgacatgaagaagtgggatcagttctacagt 748

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||||| |||| |||||||||| ||||||||||||||||||||
Db    749 gaagttctaggaagacccactacgttatctgagacaatggcaaaagctgacatatggctt 808

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || ||||||| |||| |||||||||||||| ||| |||||||||| ||||| || |||
Db    809 attcgaaactactgggattttcaatttcctcacccactcttaccaaatgttgagttcgtt 868

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||||| |||||||||| |||||||| |||
Db    869 ggaggactccactgcaaacctgccaaacccctaccgaaggaaatggaagagtttgtccag 928

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          |||||||||||||||||||||||||||||||||||||||||||||  || | ||||||| | ||
Db    929 agctctggagaaaatggtgttgtggtgttttctctggggtcgatggtcagtaacacgtca 988

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    989 gaagaaagggccaatgtaattgcatcagcccttgccaagatcccacaaaaggttctgtgg 1048

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||||| | ||||| ||||||||||||||||||||||||||
Db   1049 agatttgatgggaataaaccagatactttaggactcaatactcggctgtacaagtggata 1108

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||||
Db   1109 ccccagaatgatcttcttggtcacccaaaaaccagagcttttataactcatggtggagcc 1168

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db   1169 aatggcatctatgaggcaatctaccatggaatccctatggtgggcgttccattgtttgca 1228

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
Db   1229 gatcaacctgataacattgcacacatgaaggccaagggagcagctgttagtttggacttc 1288

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
```

```
Db   1289 cacacaatgtcgagtacagacttactcaatgcactgaagacagtaattaatgatccttta 1348

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380

Db   1349 tataaagagaatgctatgaaattatcaagaattcatcatgatcaaccagtgaagcccctt 1408

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440

Db   1409 gatcgagcagtcttctggattgaatttgtcatgcgccataaaggagccaagcaccttcgg 1468

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500

Db   1469 gttgcagcccacgacctcacctggttccagtaccactctttggatgtgactgggttcctg 1528

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560

Db   1529 ctggcctgtgtggcaactgtgatattcatcatcacaaa---atgtctgttttgtgtctgg 1585

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620

Db   1586 aagtttgttagaacaggaaagaagggaaaagagattaattacgtctgaggctggaagct 1645

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680

Db   1646 gggaaacccaataaat-gaactcctttagtttattacaacaagaa--gacgttgtgatac 1702

Qy   1681 gcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720

Db   1703 aagagattcctttcttcttgtgacaaaacatctttcaaaa 1742
```

RESULT 3
C65396
ID C65396 standard; cDNA; 1650 BP.
XX
AC C65396;
XX
DT 13-FEB-2001 (first entry)
XX
DE Human carbohydrate-modifying enzyme cDNA Incyte ID No: 2912330CB1.
XX
KW Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW autoimmune disorder; inflammatory disorder; infection; ss.
XX
OS Homo sapiens.
XX
PN WO200063351-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-US10882.
XX
PR 21-APR-1999; 99US-0130383.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX
DR WPI; 2000-672729/65.
DR P-PSDB; B28677.
XX
PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT for diagnosis, treatment, and prevention of carbohydrate metabolism
PT disorders, autoimmune/inflammatory disorders, and cancer -
XX
PS Claim 4; Page 75; 75pp; English.
XX
CC The present cDNA sequence encodes a human carbohydrate-modifying enzyme
CC (CME). CME polynucleotides and polypeptides are useful for treating and
CC diagnosing diseases associated with CME such as diabetes,
CC autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC adult respiratory distress syndrome, allergies, anaemia, asthma,
CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC complications of cancer, haemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC useful for drug screening.
XX
SQ Sequence 1650 BP; 489 A; 330 C; 354 G; 477 T; 0 other;

Query Match 71.9%; Score 1238; DB 21; Length 1650;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

```
Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60

Db     26 tgcatcaggatgtctatgaaatggacttcagctcttctcctgatacagctgagctgttac 85

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120

Db     86 tttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagccactggatg 145

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180

Db    146 aatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactgtattggca 205

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240

Db    206 tcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttgaagtttat 265

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300

Db    266 cctgtatctttaactaaaactgagtttgaggatattatcaagcagctggttaagagatgg 325

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360

Db    326 gcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaatcatgtgg 385

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420

Db    386 acatttaatgacatacttagaaagttctgtaaggatatagtttcaaataagaaacttatg 445

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480

Db    446 aagaaactacaggagtcaagatttgatgttgttcttgcagatgctgtttttccccttggt 505

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540

Db    506 gagctgctggccgagttacttaaaatacccttgtctacagcctccgcttctctcctggc 565

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600

Db    566 tacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgcctgttgtt 625

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660

Db    626 atgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatgatctatgtg 685

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720

Db    686 ctttattttgaattttggttccaaatatttgacatgaagaagtgggatcagttctacagt 745
```

```
Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          |||||| |||||||||||||||| |||| |||||||||| ||||||||||||||||||||
Db    746 gaagttctaggaagacccactacgttatctgagacaatggcaaaagctgacatatggctt 805

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          || ||||||| |||| |||||||||||||| ||| |||||||||| ||||| || |||
Db    806 attcgaaactactgggattttcaatttcctcacccactcttaccaaatgttgagttcgtt 865

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
          |||||| |||||||||||||||||||||||||||| |||||||||| |||||||| |||
Db    866 ggaggactccactgcaaacctgccaaacccctaccgaaggaaatggaagagtttgtccag 925

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          ||||||||||||||||||||||||||||||||||||||||||  || | ||||||| | ||
Db    926 agctctggagaaaatggtgttgtggtgttttctctggggtcgatggtcagtaacacgtca 985

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    986 gaagaaagggccaatgtaattgcatcagcccttgccaagatcccacaaaaggttctgtgg 1045

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| |||||||||||||| | ||||| ||||||||||||||||||||||||||||
Db   1046 agatttgatgggaataaaccagatactttaggactcaatactcggctgtacaagtggata 1105

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||| ||||| |||||||| ||||||| ||||||||| ||||||||||||
Db   1106 ccccagaatgatcttcttggtcatcccaaaaccaaagcttttatcactcatggtggaatg 1165

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttt 1200
          ||||| |||||||| || || ||||||||| |||||||||||||  ||||  | |||  |
Db   1166 aatgggatctatgaagctatttaccatggggtccctatggtgggagttcccatatttggt 1225

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          ||||| | ||||||||| |||||||||||||||| ||||||||||   | |  |||||
Db   1226 gatcagcttgataacatagctcacatgaaggccaaaggagcagctgtagaaataaacttc 1285

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          || || ||| | ||   |||  | |||| ||  |||  ||||| |||| ||| | |
Db   1286 aaaactatgacaagcgaagatttactgagggctttgagaacagtcattaccgattcctct 1345

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          ||||||||||||  ||||| |||||||||||||| |||||||||| ||||||||||
Db   1346 tataaagagaatgctatgagattatcaagaattcaccatgatcaacctgtaaagccccta 1405

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          |||||||||||||||||||| || |||||||||| |||||||||||||| ||||| |||
Db   1406 gatcgagcagtcttctggatcgagtttgtcatgcgccacaaaggagccaagcacctgcga 1465

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
             || |||||||||||||||||||||||||| || ||||| | |||||||||||||| |||
Db   1466 tcagctgcccatgacctcacctggttccagcactactctatagatgtgattgggttcctg 1525

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          ||| ||||||||||||||  |||||  |  ||||||| | || | | ||||  |
Db   1526 ctgacctgtgtggcaactgctatattcttgttcacaaaatgtttttttatttcctgtcaa 1585

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtc 1606
          || |||  || ||   | ||||  | ||| || || ||| | | ||
Db   1586 aaatttaataaaactagaaagatagaaaagagggaatagatctttc 1631


RESULT  4
V15900
ID   V15900 standard; cDNA; 2107 BP.
XX
AC   V15900;
XX
DT   26-MAY-1998  (first entry)
XX
DE   Uridine diphospho-glucuronosyltransferase 2B17 (UGT2B17) encoding cDNA.
XX
KW   Uridine diphospho-glucuronosyltransferase 2B17; UGT2B17; catalyse;
KW   androsterone; androsterone-glucuronic acid; androgen; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..51
FT                   /*tag= a
FT   CDS             52..1644
FT                   /*tag= b
FT                   /product= "UGTB17 enzyme"
FT   3'UTR           1645..2107
FT                   /*tag= c
XX
PN   WO9744466-A1.
XX
PD   27-NOV-1997.
XX
PF   16-MAY-1997;   97WO-CA00328.
XX
PR   17-MAY-1996;   96US-0649319.
XX
PA   (ENDO-) ENDORECHERCHE INC.
XX
PI   Beaulieu M,  Belanger A,  Hum DW,  Levesque E;
XX
DR   WPI; 1998-018520/02.
DR   P-PSDB; W47126.
XX
PT   DNA encoding uridine di:phospho:glucuronosyl:transferase 2B17 -
PT   which catalyses conversion of androsterone to
PT   androsterone-glucuronic acid
XX
PS   Claim 15; Pages 4-6; 53pp; English.
XX
CC   This cDNA encodes an enzyme uridine di-phosphoglucuronosyltransferase
CC   2B17 (UGT2B17). This novel enzyme catalyses the conversion of
CC   androsterone to androsterone-glucuronic acid. The UGT2B17 can be used to
CC   detect anti-UGT2B17 antibodies. The antibody can be used to detect a
CC   localised concentration of UGT2B17 or an alteration in androgen activity.
CC   The UGT2B17 can also be used to alter the concentration of an androgenic
CC   compound in a tissue, specifically dihydrotestosterone. An isolated
CC   nucleotide sequence comprising at least 30 consecutive nucleotides from
CC   the coding region of the 2107 base pair sequence, or its complement can
CC   be used to block the synthesis of UGT2B17, e.g. an expression disrupting
CC   sense or antisense fragment, or as a probe for a UGT2B17 coding sequence.
XX
SQ   Sequence 2107 BP; 657 A; 382 C; 424 G; 644 T; 0 other;


  Query Match             70.8%;  Score 1219.2;  DB 19;  Length 2107;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 1430;  Conservative    0;  Mismatches  268;  Indels   10;  Gaps    4;

Qy      4 accaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttacttt 63
          ||||||||| |||||||||||  |||||  |||||||||| || ||||||||||||||
Db     46 accaggatgtctctgaaatggatgtcagtctttctgctgatgcagctcagttgttacttt 105

Qy     64 agctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaat 123
          |||||||||||||||||||| |||||||||||| || ||||||||||||||||||| |||
Db    106 agctctgggagttgtggaaaggtgctggtgtggcccacagaatacagccattggataaat 165

Qy    124 atgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatct 183
          |||||||||||||| |||||||||||||||| |||||||||||| |||  || |||||
Db    166 atgaagacaatcctggaagagcttgttcagaggggtcatgaggtgattgtgttgacatct 225

Qy    184 tcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcct 243
          || ||||| |||||| |  || ||| | |  ||||  || ||||||| ||||||||||||
Db    226 tcggcttctattcttgtcaatgccagtaaatcatctgctattaaattagaagtttatcct 285

Qy    244 acatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatggtca 303
          |||||||||||||||| ||| || ||  || |  | ||| ||  | |  | |||||| ||
Db    286 acatctttaactaaaaatgatttggaagatttttttatgaaaatgttcgatagatggaca 345

Qy    304 ga---cattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
```

```
Db    346 tatagtatttcaaaaaatacattttggtcatatttttcacaactacaagaattgtgttgg 405

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
Db    406 gaatattctgactataatataaagctctgtgaagatgcagttttgaacaagaaacttatg 465

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
Db    466 agaaaactacaagagtcaaaatttgatgtccttctggcagatgccgttaatccctgtggt 525

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
Db    526 gagctgctggctgaactacttaacataccctttctgtacagtctccgcttctctgttggc 585

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
Db    586 tacacagttgagaagaatggtggaggatttctgttccctccttcctatgtacctgttgtt 645

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
Db    646 atgtcagaattaagtgatcaaatgattttcatggagaggataaaaaatatgatatatatg 705

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
Db    706 ctttattttgacttttggtttcaagcatatgatctgaagaagtgggaccagttttatagt 765

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
Db    766 gaagttctaggaagacccactacattatttgagacaatggggaaagctgaaatgtggctc 825

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
Db    826 attcgaacctattgggattttgaatttcctcgcccattcttaccaaatgttgattttgtt 885

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
Db    886 ggaggacttcactgtaaaccagccaaacccttgcctaaggaaatggaagagtttgtgcag 945

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
Db    946 agctctggagaaaatggtattgtggtgttttctctggggtcgatgatcagtaacatgtca 1005

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
Db   1006 gaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaaaaggttctatgg 1065

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
Db   1066 agatttgatggcaagaagccaaatactttaggttccaatactcgactgtataagtggtta 1125

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
Db   1126 ccccagaatgaccttcttggtcatcccaaaaccaaagcttttataactcatggtggaacc 1185

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
Db   1186 aatggcatctatgaggcgatctaccatgggatccctatggtgggcattcccttgtttgcg 1245

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
Db   1246 gatcaacatgataacattgctcacatgaaagccaagggagcagccctcagtgtggacatc 1305

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
Db   1306 aggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcattaatgaccctatc 1365

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
Db   1366 tataaagagaatatcatgaaattatcaagaattcatcatgatcaaccggtgaagcccctg 1425

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
Db   1426 gatcgagcagtcttctggattgagtttgtcatgcgccataaaggagccaagcaccttcgg 1485

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
Db   1486 gtcgcagcccacaacctcacctggatccagtaccactctttggatgtgatagcattcctg 1545

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
Db   1546 ctggcctgcgtggcaactatgatatttatgatcacaaaatgttgcctgttttgtttccga 1605

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
Db   1606 aagcttgccaaaacaggaaagaagaagaaaagggattagttatatcaaaagcctgaag-t 1664

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
Db   1665 ggaatgaccaaaagatgggactcctcc--tttattccagcatggagg----gttttaaat 1718

Qy   1681 gcaagatttctttcttcctgtgacaaaa 1708
Db   1719 ggaggatttccttttcctgcgacaaaa 1746
```

RESULT 5
Z95206
ID Z95206 standard; DNA; 1976 BP.
XX
AC Z95206;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B15 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW drug interaction; detect; human; single nucleotide polymorphism;
KW SNPs; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; Y78935.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 56-59; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B15
CC (UGT2B15) gene. UDP-glucuronosyltransferase (UGTs) are a family of
CC enzymes that catalyse the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC substrate in an individual. The nucleic acid molecules comprising a human
CC UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of UGT2B
CC substrate, potential drug-drug interactions and adverse side effects. The

```
CC   polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC   detecting genetic linkage related to phenotypic variation in activity or
CC   expression of UGT2B protein. The polymorphism containing nucleic acid
CC   molecules may also be used for generating genetically modified non-human
CC   animals and for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 1976 BP; 594 A; 368 C; 419 G; 595 T; 0 other;


  Query Match             70.5%;  Score 1213.6;  DB 21;  Length 1976;
  Best Local Similarity   83.1%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches  279;  Indels   10;  Gaps    3;

Qy      4 accaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttacttt 63
          ||||||||| |||||||||||||| |||||  |||||||||||||| ||||||||||||||||
Db      5 accaggatgtctctgaaatggacgtcagtctttctgctgatacagctcagttgttacttt 64

Qy     64 agctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaat 123
          |||||||| || |||||||| ||||| |||||| || ||||||||||||||||||| |||
Db     65 agctctggaagctgtggaaaggtgctagtgtggcccacagaatacagccattggataaat 124

Qy    124 atgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatct 183
          |||||||||||||| |||||||||||||||||| ||||||||||||||||||  || |||||
Db    125 atgaagacaatcctggaagagcttgttcagaggggtcatgaggtgactgtgttgacatct 184

Qy    184 tcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcct 243
          || ||||| | |||| |  || ||| | |  |||| || |||||||| ||||||||||||
Db    185 tcggcttctactcttgtcaatgccagtaaatcatctgctattaaattagaagtttatcct 244

Qy    244 acatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg--- 300
          |||||||||||||||| ||| || ||  ||    |  || ||    |  | ||||||
Db    245 acatctttaactaaaaatgatttggaagattctcttctgaaaattctcgatagatggata 304

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          |  |   ||  |||| |||  ||||||| ||||||||||||  |||||||  |    |||
Db    305 tatggtgtttcaaaaaatacattttggtcatatttttcacaattacaagaattgtgttgg 364

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
          ||||  ||||||     ||  ||  |||||||||||| |||||  ||||||||| |||||
Db    365 gaatattatgactacagtaacaagctctgtaaagatgcagttttgaataagaaacttatg 424

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          |  ||||||||||||||||  |||||  || || | ||||||||  ||  ||||||||||
Db    425 atgaaactacaagagtcaaagtttgatgtcattctggcagatgcccttaatccctgtggt 484

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||  |||||||  ||| |||||||||  ||| ||||||||| || ||  ||  ||||
Db    485 gagctactggctgaactatttaacataccctttctgtacagtcttcgattctctgttggc 544

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||||||  ||||  | | |  |||||||| |  | ||||||||||||||  |||||  |||||
Db    545 tacacatttgagaagaatggtggaggatttctgttccctccttcctatgtacctgttgtt 604

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          ||||||  |||||||||||||||||  ||||||||||||  ||||||||||||  || ||
Db    605 atgtcagaattaagtgatcaaatgattttcatggagaggataaaaaatatgatacatatg 664

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          ||||||||||||||||||  |||||  | ||||  ||||||||||||  |||||||  |||
Db    665 ctttattttgacttttggtttcaaatttatgatctgaagaagtgggaccagttttatagt 724

Qy    721 gaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctt 780
          ||||||  |||||||||||||||  |||||||||||||||||  ||||||||  || |||||
Db    725 gaagttctaggaagacccactacattatttgagacaatggggaaagctgaaatgtggctc 784

Qy    781 atgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgtt 840
          ||  |||| ||  |||  |||| |||||||||  |||||||||||||| ||||||||||||
Db    785 attcgaacctattgggattttgaatttcctcgcccattcttaccaaatgttgattttgtt 844

Qy    841 ggaggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacag 900
          ||||||  |  |||||  |||||  |||||||||||  ||||||||||||||  ||||||| |||
Db    845 ggaggacttcactgtaaaccagccaaacccctgcctaaggaaatggaagagtttgtgcag 904

Qy    901 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgaca 960
          |||||||||||||||||| ||||||||||||||||||||||  |||| |||||||||  ||
Db    905 agctctggagaaaatggtattgtggtgttttctctggggtcgatgatcagtaacatgtca 964

Qy    961 gcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtgg 1020
          | |||||| |||||  | |||||| |||||||||| |||||||||||||||||||| |||
Db    965 gaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaaaaggttctatgg 1024

Qy   1021 agatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggata 1080
          |||||||| || || || ||| || | |||||||  |||||||| ||||||||||| ||
Db   1025 agatttgatggcaagaagccaaatactttaggttccaatactcgactgtacaagtggtta 1084

Qy   1081 ccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagcc 1140
          |||||||||||||||| |||||||| |||||||  |||||||||||||||||||||||| ||
Db   1085 ccccagaatgaccttcttggtcatcccaaaaccaaagcttttataactcatggtggaacc 1144

Qy   1141 aatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgttttt 1200
          |||||||||||||||| |||||||||||||||||||||||||||||||||| ||||||
Db   1145 aatggcatctatgaggcgatctaccatgggatccctatggtgggcattcccttgtttgcg 1204

Qy   1201 gatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc 1260
          ||||||| |||||||||||||||||||| |||||||||||||  | ||  ||||| ||
Db   1205 gatcaacatgataacattgctcacatgaaagccaagggagcagccctcagtgtggacatc 1264

Qy   1261 aacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttta 1320
          |  || ||||| |||| |||  |||| |||||| |||| |||| ||||||||| ||| |
Db   1265 aggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcattaatgaccctgtc 1324

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagccccctg 1380
          |||||||||||| | |||||||||||||||||||| |||| |||||| | |||||||||
Db   1325 tataaagagaatgtcatgaaattatcaagaattcatcatgaccaaccaatgaagcccctg 1384

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          |||||||||||||||||||||| |||||||||| ||||||||||||||| |||||||||
Db   1385 gatcgagcagtcttctggattgagtttgtcatgcgccacaaaggagccaagcaccttcga 1444

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
          || |||||  ||  |||||||||||| ||||||||||||||||||||||||||  ||  |||
Db   1445 gtcgcagctcacaacctcacctggatccagtaccactctttggatgtgatagcattcctg 1504

Qy   1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
          |||||||| |||||||||||||||||||||||||||| ||||| |||||||||||| |
Db   1505 ctggcctgcgtggcaactgtgatatttatcatcacaaaattttgcctgttttgtttccga 1564

Qy   1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
          ||| |||| | || ||| ||||||   ||||||||||||||| ||  |    ||||| |
Db   1565 aagcttgccaaaacaggaaagaagaagaaaagagattagttatatcaaaagcctgaag-t 1623

Qy   1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
          ||||   | || |||| ||||  || |  | | |       || | | |  | ||  ||
Db   1624 ggaatgactgaaagatgggactcctcctttattt------cagcatggagggttttaaat 1677

Qy   1681 gcaagatttctttcttcctgtgacaaaaaa 1710
          | | |||||| || |||||||||||||| |
Db   1678 ggaggatttccttttttcctgtgacaaaaca 1707


RESULT   6
Z95201
ID   Z95201 standard; DNA; 1686 BP.
XX
AC   Z95201;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B7 exon 1 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
```

XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 2; Page 45; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 1 of the human
CC UDP-glucuronosyltransferase 2B7 (UGT2B7) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,
CC also to predict their rate of metabolism of UGT2B substrate, potential
CC drug-drug interactions and adverse side effects. The polymorphisms can be
CC used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC linkage related to phenotypic variation in activity or expression of
CC UGT2B protein. The polymorphism containing nucleic acid molecules may
CC also be used for generating genetically modified non-human animals and
CC for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 1686 BP; 548 A; 286 C; 295 G; 557 T; 0 other;

Query Match 35.3%; Score 608.2; DB 21; Length 1686;
Best Local Similarity 89.4%; Pred. No. 2.6e-150;
Matches 655; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          ||||||||||| || ||||||||||||||||  || |||| ||||| || || | || |
Db    397 tgcaccaggatgtctgtgaaatggacttcagtaattttgctaatacaactgagcttttgc 456

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
          |||||||||||| ||||||||| |||||||||||||| ||||||||||||||||||||||
Db    457 tttagctctgggaattgtggaaaggtgctggtgtgggcagcagaatacagccattggatg 516

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          ||||| |||||||||||| | |||||| ||||||||||||||||||||||||||||||||
Db    517 aatataaagacaatcctggatgagcttattcagagaggtcatgaggtgactgtactggca 576

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          |||||||||||||||||||||||||||  |  ||||| |||||||| ||||| |||||
Db    577 tcttcagcttccattctttttgatcccaacaactcatccgctcttaaaattgaaatttat 636

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
          || |||||||||||||||||| || |||||| |||||||||||||| ||||||||||||
Db    637 cccacatctttaactaaaactgagttggagaatttcatcatgcaacagattaagagatgg 696

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          |||||| ||| |||||||| ||||||||||||||||||||| ||| |||||| |||
Db    697 tcagaccttccaaaagatacattttggttatatttttcacaagtacaggaaatcatgtca 756

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
             |||   |||||||  |||||| ||||||||||||||||||||||||||||||| |||||
Db    757 atatttggtgacataactagaaagttctgtaaagatgtagtttcaaataagaaatttatg 816
```

```
Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          |||||| |||||||||||||||||||| || |||||||||||||| |||||||||| ||
Db    817 aaaaaagtacaagagtcaagatttgacgtcatttttgcagatgctatttttccctgtagt 876

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||||||||||| |||| ||||||||| ||||||||||||||| |||| ||||||||
Db    877 gagctgctggctgagctatttaacataccctttgtgtacagtctcagcttctctcctggc 936

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          |||||  |||||| ||| |||||||||| | ||||||||||||||||||| ||||| |||||
Db    937 tacacttttgaaaagcatagtggaggatttattttccctccttcctacgtacctgttgtt 996

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          |||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    997 atgtcagaattaactgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 1056

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          ||||| |||||||||||||| |||| | ||| ||||||||||||||||||||||| |||
Db   1057 ctttactttgacttttggttcgaaatatttgacatgaagaagtgggatcagttttatagt 1116

Qy    721 gaagttttaggaa 733
          |||||| |||| |
Db   1117 gaagttctaggta 1129
```

RESULT 7
Z95193
ID Z95193 standard; DNA; 1323 BP.
XX
AC Z95193;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UGT2B4 exon 1 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 1; Page 31; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 1 of the human
CC UDP-glucuronosyltransferase 2B4 (UGT2B4) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,

```
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 1323 BP; 403 A; 225 C; 253 G; 442 T; 0 other;


  Query Match             34.0%;  Score 584.8;  DB 21;  Length 1323;
  Best Local Similarity   85.8%;  Pred. No. 3.3e-144;
  Matches  649;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
           |||| |||||||  || ||||||||||||||| ||||||||||||||| || ||  ||||||
Db     168 tgcatcaggatgtctatgaaatggacttcagctcttctgctgatacagctgagctgttac 227

Qy      61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
           |||||||||||||||||||||||| |||||||||||  ||  ||||||  ||||||  ||||||
Db     228 tttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagccactggatg 287

Qy     121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
           |||||  |||||||||||| | ||  |||||  |||||||||||||||||||||||||  ||||||
Db     288 aatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactgtattggca 347

Qy     181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
           ||||||||||||||  ||| ||||||||     ||||  ||||||||||||||||||||
Db     348 tcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttgaagtttat 407

Qy     241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
           |||  |||||||||||||||||| ||||||  |||| |||| |||  | ||||||||||||||
Db     408 cctgtatctttaactaaaactgagtttgaggatattatcaagcagctggttaagagatgg 467

Qy     301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
            ||||  ||| ||||||  |  |||||||  |||||||||||||| ||||||||||  ||||||
Db     468 gcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaatcatgtgg 527

Qy     361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
             |||  |||||||| |||||||  |||||||| ||| ||||||||||||||||  ||||||
Db     528 acatttaatgacatacttagaaagttctgtaaggatatagtttcaaataagaaacttatg 587

Qy     421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
           ||  ||||||||  ||||||||||||||  | ||| ||||||||||||||||  ||||  ||||
Db     588 aagaaactacaggagtcaagatttgatgttgttcttgcagatgctgttttcccctttggt 647

Qy     481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
           ||||||||||  | |  |||||||  ||||  |||||  ||||||||  ||||||||
Db     648 gagctgctggccgagttacttaaaatacccttttgtctacagcctccgcttctctcctggc 707

Qy     541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
           |||  ||||||||| ||| ||||||||||  | ||||||||||||  | ||| |||||
Db     708 tacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgcctgttgtt 767

Qy     601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
           ||||||  ||  |||||||  |||||||||||||| ||||||||||||||||||||||||||||
Db     768 atgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatgatctatgtg 827

Qy     661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
           ||||||||||  |||||||||||||| | ||| ||||||||||||||||||||  ||||||
Db     828 ctttattttgaattttggttccaaatatttgacatgaagaagtgggatcagttctacagt 887

Qy     721 gaagttttaggaagacccactaccttatttgagaca 756
           ||||||  |||| |       |   | ||| |  |||
Db     888 gaagttctaggtaagtaacttttttgattggtaaca 923


RESULT   8
Z95207
ID   Z95207 standard; DNA; 2312 BP.
XX
AC   Z95207;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B15 exon 1 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism;
KW   SNPs; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions -
XX
PS   Example 3; Page 60-61; 72pp; English.
XX
CC   This sequence represents the nucleotide sequence of exon 1 of the human
CC   UDP-glucuronosyltransferase 2B15 (UGT2B15) gene.
CC   UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC   the glucuronic acid conjugation of a wide range of endogenous and
CC   exogenous substrates. The UGT2B gene subfamily encode steroid
CC   metabolizing isoforms in the liver. Alteration of the expression or
CC   function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC   can be used to detect altered UGT2B metabolism of a substrate in an
CC   individual. The nucleic acid molecules comprising a human UGT2B sequence
CC   polymorphism can be used in screening assays for genotyping individuals,
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 2312 BP; 765 A; 360 C; 414 G; 773 T; 0 other;


  Query Match             29.0%;  Score 498.8;  DB 21;  Length 2312;
  Best Local Similarity   79.6%;  Pred. No. 1.8e-121;
  Matches  603;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy       4 accaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttacttt 63
           |||||||||  |||||||||||||  |||||  ||||||||||||  ||||||||||||||
Db     696 accaggatgtctctgaaatggacgtcagtctttctgctgatacagctcagttgttacttt 755

Qy      64 agctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaat 123
           ||||||||  ||  ||||||||  |||||  ||||||  ||  ||||||||||||||||  |||
Db     756 agctctggaagctgtggaaaggtgctagtgtggcccacagaatacagccattggataaat 815

Qy     124 atgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatct 183
           ||||||||||||||  ||||||||||||||||  ||||||||||||||||  ||  |||||
Db     816 atgaagacaatcctggaagagcttgttcagaggggtcatgaggtgactgtgttgacatct 875

Qy     184 tcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcct 243
           ||  |||||  | ||||  |  ||  |||  | |  ||||  ||  |||||||  ||||||||||||
Db     876 tcggcttctactcttgtcaatgccagtaaatcatctgctattaaattagaagtttatcct 935
```

```
Qy    244 acatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg--- 300
          |||||||||||||||| ||| || ||  ||    |  || ||    |  | ||||||
Db    936 acatctttaactaaaaatgatttggaagattctcttctgaaaattctcgatagatggata 995

Qy    301 tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgg 360
          |  |   ||  |||| |||  ||||||| ||||||||||||  |||||||  |    |||
Db    996 tatggtgtttcaaaaaatacattttggtcatatttttcacaattacaagaattgtgttgg 1055

Qy    361 gaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatg 420
          ||||  ||||||     ||  ||  ||||||||||| |||||  ||||||||| |||||
Db   1056 gaatattatgactacagtaacaagctctgtaaagatgcagttttgaataagaaacttatg 1115

Qy    421 aaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggt 480
          |  ||||||||||||||||  |||||  || || | ||||||||  ||  ||||||||||
Db   1116 atgaaactacaagagtcaaagtttgatgtcattctggcagatgcccttaatccctgtggt 1175

Qy    481 gagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggc 540
          |||||  |||||||  |||  |||||||||  |||  |||||||||| || ||  ||  ||||
Db   1176 gagctactggctgaactatttaacataccctttctgtacagtcttcgattctctgttggc 1235

Qy    541 tacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtt 600
          ||||||  ||||  | |  |  ||||||||  |  | ||||||||||||||  |||||  |||||
Db   1236 tacacatttgagaagaatggtggaggatttctgttccctccttcctatgtacctgttgtt 1295

Qy    601 atgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtg 660
          ||||||  |||||||||||||||||||  ||||||||||||  ||||||||||||||  || ||
Db   1296 atgtcagaattaagtgatcaaatgattttcatggagaggataaaaaatatgatacatatg 1355

Qy    661 ctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagt 720
          ||||||||||||||||||  |||||  |  ||||  ||||||||||||  ||||||||  |||
Db   1356 ctttattttgacttttggtttcaaatttatgatctgaagaagtgggaccagttttatagt 1415

Qy    721 gaagttttaggaagacccactaccttatttgagacaat 758
          ||||||  ||||  |   |   |  || | | | |   ||
Db   1416 gaagttctaggtaagtcatgtgtctaactggtgcttat 1453
```

```
RESULT   9
Z95205
ID   Z95205 standard; DNA; 596 BP.
XX
AC   Z95205;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B7 exon 5 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions  -
XX
PS   Example 2; Page 47-48; 72pp; English.
XX
CC   This sequence represents the nucleotide sequence of exon 5 of the human
CC   UDP-glucuronosyltransferase 2B7 (UGT2B7) gene.
CC   UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC   the glucuronic acid conjugation of a wide range of endogenous and
CC   exogenous substrates. The UGT2B gene subfamily encode steroid
CC   metabolizing isoforms in the liver. Alteration of the expression or
CC   function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC   can be used to detect altered UGT2B metabolism of a substrate in an
CC   individual. The nucleic acid molecules comprising a human UGT2B sequence
CC   polymorphism can be used in screening assays for genotyping individuals,
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 596 BP; 199 A; 97 C; 115 G; 185 T; 0 other;
```

```
  Query Match             19.9%;  Score 343;  DB 21;  Length 596;
  Best Local Similarity   92.6%;  Pred. No. 1e-80;
  Matches  373;  Conservative    0;  Mismatches   25;  Indels    5;  Gaps    1;

Qy   1320 atataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccct 1379
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||  ||||||||
Db     20 atataaagagaatgttatgaaattatcaagaattcaacatgatcaaccagtgaagcccct 79

Qy   1380 ggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcg 1439
          |||||||||||||||||||||||||||||||||||| ||||||||||| ||||||||||
Db     80 ggatcgagcagtcttctggattgaatttgtcatgcgccacaaaggagctaaacaccttcg 139

Qy   1440 agttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttct 1499
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||
Db    140 ggttgcagcccacgacctcacctggttccagtaccactctttggatgtgattgggttcct 199

Qy   1500 gctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1559
          |||||  |||||||||||||||||||||||||||  |||||||  | ||||||||||||||||
Db    200 gctggtctgtgtggcaactgtgatatttatcgtcacaaaatgttgtctgttttgtttctg 259

Qy   1560 gaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagc 1619
          ||||||||||||||||  ||||||||||||  ||||||||| |||||  |||||||||
Db    260 gaagtttgctagaaaagcaaagaagggaaaaaatgattagttatatctgagatttgaagc 319

Qy   1620 tggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtta 1679
          |||||||||  |||||  |  |||  ||||||||||||||||||||     ||||||||| |
Db    320 tggaaaacctgataggtgagactacttcagtttattccagcaag-----aaagattgtga 374

Qy   1680 tgcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
          |||||||||||||||||||| ||||||||||||||| ||||
Db    375 tgcaagatttctttcttcctgagacaaaaaaaaaaaaagaaaa 417
```

```
RESULT  10
Q27370
ID   Q27370 standard; cDNA; 2368 BP.
XX
AC   Q27370;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br2.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             30..801
```

```
FT                   /*tag= a
FT   polyA_signal    2347..2352
FT                   /*tag= b
FT                   /number= 1
FT   polyA_signal    2355..2360
FT                   /*tag= c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; R26154.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver billrubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (Q27369) and HUGBr2
CC   (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2368 BP; 609 A; 532 C; 566 G; 661 T; 0 other;
```

```
  Query Match           18.1%;  Score 311.4;  DB 13;  Length 2368;
  Best Local Similarity 52.2%;  Pred. No. 3.7e-72;
  Matches  796;  Conservative    0;  Mismatches  711;  Indels   18;  Gaps    4;

Qy      40 ctgatacatctcagttgttactttagctctgggagttgtggaaaagtgctggtgtgggcc 99
           ||| | |  |||    ||  |     ||  |      ||||||| ||| |||||  | ||
Db      75 ctgctgctcctcctcagtgtccagccctgggctgagagtggaaaggtgttggtggtgccc 134

Qy     100 gcagaatacagccattggatgaatatgaagacaatcctgaaagagcttgttcagagaggt 159
             | ||   |||||  ||| | |  |||  |     | ||   |||||   |   |||||
Db     135 actgatggcagcccctggctcagcatgcgggaggccttgcgggagctccatgccagaggc 194

Qy     160 catgaggtgactgtactggcatcttcagcttccattctttttgatcccaatgatgcatcc 219
           ||  ||| |   || ||  |  |   || |     |     | |     |  |    | |
Db     195 caccaggcggtggtcctcaccccaagaggtgaatatgcacatcaagaagagaaatttttc 254

Qy     220 actcttaaatttgaagtttatcctacatctttaactaaaactgaatttgagaatatcatc 279
           || || |       ||  ||| ||    | |  ||  | |  |||||||||     | |
Db     255 accctga------cagcctatgctgttccatggacccagaaggaatttgatcgcgttacg 308

Qy     280 atgcaacaggttaagagatggtcagacattcgaaaagatagctttggttatatttttca 339
             ||    |  | |  | |       | ||  || |||     ||  |      | |||
Db     309 ctgggctacactcaagggt------tctttgaaacagaacatcttctgaagagatattct 362

Qy     340 caagaacaagaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgta 399
             |      | |||  ||     |  ||| |        ||   |  |  |||   ||  ||
Db     363 agaagtatggcaattatgaacaatgtatctttggcccttcataggtgttgtgtggagcta 422

Qy     400 gtttcaaataagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgca 459
             |    ||| ||    | ||  |   | ||  | |   |    |||||  | |||||  ||

Db     423 ctgcataatgaggccctgatcaggcacctgaatgctacttcctttgatgtggttttaaca 482

Qy     460 gatgctgtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtac 519
           ||  | |||   | ||| || | | ||||||||  |   ||    || |    |||||
Db     483 gaccccgttaacctctgcggggcggtgctggctaagtacctgtcgattcctgctgtgttt 542

Qy     520 agtctccgctttactcctggctacacaattgaaaggcacagtggaggactgatttttccct 579
             |    | |  | |||  |  ||  |       | |  ||       | |  | |   |||
Db     543 ttttggaggtacattccatgtgacttagactttaagggcacacagtgtccaaat---cct 599

Qy     580 ccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagg 639
           | |||||  || ||||   |  |  | |        || || |||| ||| || | |||
Db     600 tcctcctatattcctaagttactaacgaccaattcagaccacatgacattcctgcaaagg 659

Qy     640 gtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatatgaag 699
           || || || ||| ||||    ||      |   ||      | ||| |  | |  |
Db     660 gtcaagaacatgctctaccctctggccctgtcctacatttgccatacttttttctgcccct 719

Qy     700 aagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagacaatg 759
           |      ||   ||  |  |||  ||||    |||      ||  |   ||   | ||
Db     720 tatgcaagtc---ttgcctctgagctttttcagagagaggtgactgtccaggacctattg 776

Qy     760 ggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatccattc 819
           |    ||      | |||||  |  |||    ||          | |  |||   ||  ||
Db     777 agctctgcatctgtctggctgtttagaagtgactttgtgaaggattaccctaggcccatc 836

Qy     820 ttaccaaacgttgattttgttggaggattccactgcaaacctgccaaacccctacctaag 879
           | || ||  | |  ||  |||| ||  || ||||    |       || ||       ||
Db     837 atgcccaatatggtcttcattgggggcatcaactgttgccaacgggaagccactatctag 896

Qy     880 gaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgggg 939
           ||| | || |   |   | |     |||||||| ||||  |||||| || ||| |||
Db     897 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggaa 956

Qy     940 tcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaag 999
           ||| || |     | ||  |||   | |  || |  | ||||||    ||  | | |||
Db     957 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 1016

Qy    1000 atcccacaaaaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaat 1059
           ||||| || |  || |||||| | |     || |   ||||       | |    |||
Db    1017 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 1076

Qy    1060 actcggctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagct 1119
           ||    ||   |||||| |||||| || || || || ||||| || |  ||| | ||
Db    1077 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 1136

Qy    1120 tttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatg 1179
           ||||| || |||| |||  || ||||  | |||||     || | | ||||  | || |||
Db    1137 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 1196

Qy    1180 gtgggcattccattgttttttgatcaacctgataacattgctcacatgaaggccaaggga 1239
           |||   || || ||||||  ||||||    || ||       | |||| || | ||||||
Db    1197 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 1256

Qy    1240 gcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaag 1299
           || |  || |   || | |     |||| |   |  ||| |    ||||| |  ||
Db    1257 gctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctcaaaaa 1316

Qy    1300 acagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacat 1359
           |||| || |||||       || || ||||| || |||    | || ||  ||||  |
Db    1317 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 1376

Qy    1360 gatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccac 1419
           || |  || ||  |||| ||||| |  || || ||||||  | || ||||| |||   |||
Db    1377 gaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 1436

Qy    1420 aaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactct 1479
           || || ||   ||||||  ||    |||||||| ||||||||||| |||||||||  ||
Db    1437 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 1496

Qy    1480 ttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaag 1539
           ||||| ||||||||  || ||  |||||    ||   || |||   || |||| |   ||
Db    1497 ttggacgtgattggtttcctcttggccgtgctgctgacagtggccttcatcacctttaaa 1556
```

```
Qy   1540 ttttgtctgttttgtttctggaagt 1564
          | ||||   | | | | | |||| |
Db   1557 tgttgtgcttatggctaccggaaat 1581


RESULT 11
Q27369
ID   Q27369 standard; cDNA; 2351 BP.
XX
AC   Q27369;
XX
DT   27-JAN-1993 (first entry)
XX
DE   HUG-Br1.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             16..784
FT                   /*tag= a
FT   polyA_signal    2330..2335
FT                   /*tag= b
FT                   /number= 1
FT   polyA_signal    2338..2343
FT                   /*tag= c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS, Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; R26153.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated. They are referred to as HUGBr1 (Q27369) and HUGBr2
CC   (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;


  Query Match          17.6%;  Score 302.6;  DB 13;  Length 2351;
  Best Local Similarity  57.3%;  Pred. No. 7.7e-70;
  Matches  567;  Conservative    0;  Mismatches  419;  Indels    3;  Gaps    1;

Qy    576 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 635
          |||    |||||| | || |    | | ||         ||||| ||||| ||| || |
Db    579 cccattctcctactacgtgcccaggcctctctcctctcattcagatcacatgaccttcctgca 638

Qy    636 gagggtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 695
          | |||| || || ||| ||  ||   |||      |||||  || | |  ||  | ||
Db    639 gcgggtgaagaacatgctcattgccttttcacagaactttctgtgcgacgtggtttattc 698

Qy    696 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 755
            | |       |   ||  |  ||| |  |    |||      ||  |    ||
Db    699 cccgtatgcaaccc---ttgcctcagaattccttcagagagaggtgactgtccaggacct 755

Qy    756 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 815
          | || |    ||     | |||||  |  |||    ||         | |  |||   ||
Db    756 attgagctctgcatctgtctggctgtttagaagtgactttgtgaaggattaccctaggcc 815

Qy    816 attcttaccaaacgttgattttgttggaggattccactgcaaacctgccaaaccctacc 875
            || | || ||  | | |||||||||| ||| || |||||    |     || || ||| |
Db    816 catcatgcccaatatggtttttgttggtggaatcaactgccttcaccaaaatccactatc 875

Qy    876 taaggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctct 935
            ||||| | || |  |   |  |    ||||||||| ||||  ||||||| || ||| |
Db    876 ccaggaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctcttt 935

Qy    936 ggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgc 995
          ||  ||| || |     | ||  |||   | |  || |  | ||||||    ||  | |
Db    936 ggaatcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttggg 995

Qy    996 caagatcccacaaaaggttctgtggagatttgacgggaataaaccagatgccttaggtct 1055
          ||| ||||| || |  || |||||| | |     || |    ||||       | |
Db    996 caaaatccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaa 1055

Qy   1056 caatactcggctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccag 1115
          ||| ||    ||    |||||| ||||||| || || || || ||||| || |  ||| |
Db   1056 caacacgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccg 1115

Qy   1116 agcttttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccc 1175
           || ||||| || |||| |||  || ||||  | |||||    || | | |||| | ||
Db   1116 tgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcc 1175

Qy   1176 tatggtgggcattccattgttttttgatcaacctgataacattgctcacatgaaggccaa 1235
          ||||||   || || ||||||  ||||||    || ||       | |||| || | ||
Db   1176 catggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaa 1235

Qy   1236 gggagcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcact 1295
          ||||||  | || |   || |  |      |||| |   |  |||  |    ||||| |
Db   1236 gggagctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctca 1295

Qy   1296 gaagacagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattca 1355
           ||  |||| || |||||       || || ||||| || |||    | || ||  ||||
Db   1296 aaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttca 1355

Qy   1356 acatgatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgcc 1415
           | || |  || ||  |||| ||||| |  || || ||||||  | || |||||  |||
Db   1356 caaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgag 1415

Qy   1416 ccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtacca 1475
           ||||| || ||   |||||| ||    |||||||| |||||||||||||| |||||||||
Db   1416 gcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtacca 1475

Qy   1476 ctctttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcac 1535
           || ||||| |||||||| || ||  |||||    ||   || |||   || |||| |
Db   1476 ttccttggacgtgattggtttcctcttggccgtgctgctgacagtggccttcatcacctt 1535

Qy   1536 aaagttttgtctgttttgtttctggaaagt 1564
           || | ||||   | | | | | |||| |
Db   1536 taaatgttgtgcttatggctaccggaaat 1564


RESULT 12
Z95198
ID   Z95198 standard; DNA; 1589 BP.
XX
AC   Z95198;
XX
DT   05-JUN-2000 (first entry)
```

XX
DE Human UGT2B4 exon 6 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 1; Page 33-34; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 6 of the human
CC UDP-glucuronosyltransferase 2B4 (UGT2B4) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,
CC also to predict their rate of metabolism of UGT2B substrate, potential
CC drug-drug interactions and adverse side effects. The polymorphisms can be
CC used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC linkage related to phenotypic variation in activity or expression of
CC UGT2B protein. The polymorphism containing nucleic acid molecules may
CC also be used for generating genetically modified non-human animals and
CC for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 1589 BP; 467 A; 312 C; 293 G; 517 T; 0 other;

Query Match 15.7%; Score 271; DB 21; Length 1589;
Best Local Similarity 84.8%; Pred. No. 1.4e-61;
Matches 340; Conservative 0; Mismatches 55; Indels 6; Gaps 3;

```
Qy   1320 atataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccct 1379
          ||||||||||||||  |||||||||||||||||||||| ||||||||||||||| ||||||||
Db    731 atataaagagaatgctatgaaattatcaagaattcatcatgatcaaccagtgaagcccct 790

Qy   1380 ggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcg 1439
          |||||||||||||||||||||||||||||||||| ||| |||||||||||| ||||||||
Db    791 tgatcgagcagtcttctggattgaatttgtcatgcgccataaaggagccaagcaccttcg 850

Qy   1440 agttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttct 1499
          |||||||||||| |||||||||||||||||||||||||||||||||||||| |||||| ||
Db    851 ggttgcagcccacgacctcacctggttccagtaccactctttggatgtgactgggttcct 910

Qy   1500 gctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1559
          |||||||||||||||||||||||||| ||||||||||    |||||||||||| ||||
Db    911 gctggcctgtgtggcaactgtgatattcatcatcacaaa---atgtctgttttgtgtctg 967

Qy   1560 gaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagc 1619
          |||||||| ||||| ||| |||||||| |||||||||| ||| ||||||   | |||||
Db    968 gaagtttgttagaacaggaaagaaggggaaaagagattaattacgtctgaggctggaagc 1027

Qy   1620 tggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtta 1679
          ||| ||||| ||| || | ||  ||| |||||||| || ||||||   |    || ||
Db   1028 tgggaaacccaataaat-gaactcctttagtttattacaacaagaa--gacgttgtgata 1084

Qy   1680 tgcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaa 1720
               ||||| |||||||| |||||||||| |      ||||
Db   1085 caagagattcctttcttcttgtgacaaaacatctttcaaaa 1125
```

RESULT 13
Z45118
ID Z45118 standard; DNA; 735 BP.
XX
AC Z45118;
XX
DT 28-FEB-2000 (first entry)
XX
DE UDP-glucuronosyltransferase 1 (UGT1) exons 2-5 nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnose; ss.
XX
OS Homo sapiens.
XX
PN WO9957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99WO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;
XX
DR WPI; 2000-052981/04.
DR P-PSDB; Y57100.
XX
PT New nucleic acid representing polymorphisms in the human uridine
PT diphosphate glucuronosyltransferase gene, used for diagnosis and
PT evaluation of drug metabolism -
XX
PS Examples; Page 43-44; 63pp; English.
XX
CC Nucleotide sequences Z45110-Z45118 are exons 1A-1J of the human uridine
CC diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a family
CC of enzymes that catalyse the glucuronic acid conjugation of a wide range
CC of endogenous and exogenous substrates including phenols, alcohols,
CC amines and fatty acids. Many of the reactions catalysed by UGTs result in
CC toxic substances being converted to compounds which are more water
CC soluble and are excreted. The invention relates to and identifies UGT1
CC polymorphisms (Z45004-Z45041). The polymorphism sequences are useful as
CC probes for detecting UGT1 locus polymorphisms, indicative of altered UGT1
CC expression or activity. These polymorphisms are associated with
CC Crigler-Najjar and Gilbert syndromes (unconjugated hyperbilirubinaemia)
CC and drug metabolism. The genotyping of the UGT1 gene is used to predict
CC the rate of metabolism of UGT1 substrates, possible drug-drug
CC interactions and adverse side effects (i.e. to optimize drug dosage), and
CC to screen for diseases caused by exposure to toxins and to study the
CC effects of polymorphisms on enzymatic activity. The UGT1 sequences,
CC including polymorphisms, can also be used to produce the corresponding
CC protein (or its fragments) or to generate transgenic animals or modified
CC cells e.g. for pharmacogenetic screening.
XX
SQ Sequence 735 BP; 189 A; 179 C; 187 G; 180 T; 0 other;

Query Match 15.2%; Score 262.6; DB 21; Length 735;

```
  Best Local Similarity   61.5%;  Pred. No. 1.6e-59;
  Matches  421;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy     880 gaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgggg 939
           ||| | || |  |   |  |    |||||||| ||||  ||||||| || ||| ||||
Db       1 gaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctttggga 60

Qy     940 tcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaag 999
           ||| || |     | ||  |||   | |  || |  | ||||||    ||  | | |||
Db      61 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 120

Qy    1000 atcccacaaaaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaat 1059
           ||||| || |  || |||||| | |     || |    ||||        | |    |||
Db     121 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 180

Qy    1060 actcggctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagct 1119
           ||    ||    |||||| ||||||| || || || || ||||| || |  ||| | ||
Db     181 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 240

Qy    1120 tttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatg 1179
           ||||| || |||| |||  || ||||  | |||||    || | | ||||  | || |||
Db     241 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 300

Qy    1180 gtgggcattccattgttttttgatcaacctgataacattgctcacatgaaggccaaggga 1239
           |||   || || ||||||  ||||||    || ||       | |||| || | ||||||
Db     301 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 360

Qy    1240 gcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaag 1299
           || |  || |   || |  |      |||| |   |  |||  |    ||||| || ||
Db     361 gctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 420

Qy    1300 acagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacat 1359
            |||| || |||||       || || ||||| || |||    | || ||  ||||  |
Db     421 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 480

Qy    1360 gatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccac 1419
           || |  || ||  |||| ||||| |  || || |||||| | || |||||  |||   |||
Db     481 gaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 540

Qy    1420 aaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactct 1479
           || || ||   |||||| ||    ||||||||| |||||||||||| |||||||||  ||
Db     541 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 600

Qy    1480 ttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaag 1539
           ||||| ||||||||  || ||  |||||   |||   || |||   || |||| |   ||
Db     601 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 660

Qy    1540 ttttgtctgttttgtttctggaagt 1564
           | ||||   | | | | | |||| |
Db     661 tgttgtgcttatggctaccggaaat 685


RESULT  14
Z95211
ID   Z95211 standard; DNA; 978 BP.
XX
AC   Z95211;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B15 exon 5 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism;
KW   SNPs; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions  -
XX
PS   Example 3; Page 62; 72pp; English.
XX
CC   This sequence represents the nucleotide sequence of exon 5 of the human
CC   UDP-glucuronosyltransferase 2B15 (UGT2B15) gene.
CC   UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC   the glucuronic acid conjugation of a wide range of endogenous and
CC   exogenous substrates. The UGT2B gene subfamily encode steroid
CC   metabolizing isoforms in the liver. Alteration of the expression or
CC   function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC   can be used to detect altered UGT2B metabolism of a substrate in an
CC   individual. The nucleic acid molecules comprising a human UGT2B sequence
CC   polymorphism can be used in screening assays for genotyping individuals,
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 978 BP; 321 A; 187 C; 162 G; 308 T; 0 other;


  Query Match             14.5%;  Score 250.4;  DB 21;  Length 978;
  Best Local Similarity   81.3%;  Pred. No. 2.9e-56;
  Matches  317;  Conservative    0;  Mismatches   66;  Indels    7;  Gaps    2;

Qy    1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagccctg 1380
           |||||||||||| | |||||||||||||||||||| ||||| |||||| | |||||||||
Db     327 tataaagagaatgtcatgaaattatcaagaattcatcatgaccaaccaatgaagccctg 386

Qy    1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
           |||||||||||||||||||||| |||||||||| |||||||||||||| |||||||||
Db     387 gatcgagcagtcttctggattgagtttgtcatgcgccacaaaggagccaagcaccttcga 446

Qy    1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           || ||||| ||  |||||||||||| ||||||||||||||||||||||||| |  || |||
Db     447 gtcgcagctcacaacctcacctggatccagtaccactctttggatgtgatagcattcctg 506

Qy    1501 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1560
           |||||||| ||||||||||||||||||||||||||||| ||||| ||||||||||| |
Db     507 ctggcctgcgtggcaactgtgatatttatcatcacaaaattttgcctgttttgtttccga 566

Qy    1561 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1620
           |||  |||| | || ||| ||||||   |||||||||||||||| ||  |    ||||| |
Db     567 aagcttgccaaaacaggaaagaagaagaaaagagattagttatatcaaaagcctgaag-t 625

Qy    1621 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1680
           ||||   | || |||| ||||  || |  | | |       || | | |  | ||   ||
Db     626 ggaatgactgaaagatgggactcctcctttattt------cagcatggagggttttaaat 679

Qy    1681 gcaagatttctttcttcctgtgacaaaaaa 1710
           | | |||||| || |||||||||||||| |
Db     680 ggaggatttccttttttcctgtgacaaaaca 709


RESULT  15
C03286
```

```
ID   C03286 standard; cDNA; 350 BP.
XX
AC   C03286;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 3284.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; G03280.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 3284; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 350 BP; 108 A; 69 C; 77 G; 96 T; 0 other;


  Query Match            13.4%;  Score 230;  DB 21;  Length 350;
  Best Local Similarity  78.6%;  Pred. No. 4.4e-51;
  Matches  275;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    769 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 828
          || |||||||| || ||||  |  |||  |||| |||||||||| |||| |  ||| |||
Db      1 gagatatggctaatacgaacatattgggattttgaatttcctcaaccataccaacctaac 60

Qy    829 gttgattttgttggaggattccactgcaaacctgccaaacccctacctaaggaaatggag 888
           |||| |||||||||||||| ||||| |||||||||||  |  | ||||||||||||||
Db     61 tttgagtttgttggaggattgcactgtaaacctgccaaagctttgcctaaggaaatggaa 120

Qy    889 gagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgata 948
           | ||||| ||||| || || ||| ||||| ||||||||||||||||||||||| || |
Db    121 aattttgtccagagttcaggggaagatggtattgtggtgttttctctggggtcactgttt 180

Qy    949 agtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaa 1008
             ||  | |||| ||||| ||| ||| | |||||  |||||||||| ||||||||||
Db    181 caaaatgttacagaagaaaaggctaatatcattgcttcagcccttgcccagatcccacag 240

Qy   1009 aaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaatactcggctg 1068
          |||||  | ||||| |   | || || ||||||    | |||||   ||||||||||||
Db    241 aaggtgttatggaggtacaaaggaaaaaaaaccatccacattaggagccaatactcggctg 300

Qy   1069 tacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagc 1118
          ||  | ||||||||||||||||| ||||| |||||||| ||||||| |||
Db    301 tatgattggataccccagaatgatcttcttggtcatcccaaaaccaaagc 350


Search completed: April 30, 2001, 19:07:57
Job time: 36891 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:33:45 ; Search time 330.07 Seconds
(without alignments)
910.915 Million cell updates/sec

Title: US-09-721-183-18
Perfect score: 1722
Sequence: 1 tgcaccaggatgactctgaa.....acaaaaaaaaaaaaaaaaaa 1722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	324.2	18.8	2339	5	PCT-US92-00282-2	Sequence 2, Appli
	2	305.8	17.8	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	3	106.6	6.2	1561	5	PCT-US92-00282-25	Sequence 25, Appl
	4	81.6	4.7	783	5	PCT-US92-00282-22	Sequence 22, Appl
	5	64.2	3.7	1008	5	PCT-US92-00282-10	Sequence 10, Appl
	6	59.4	3.4	1066	5	PCT-US92-00282-14	Sequence 14, Appl
	7	58	3.4	1197	5	PCT-US92-00282-12	Sequence 12, Appl
	8	52.8	3.1	1448	5	PCT-US92-00282-16	Sequence 16, Appl
	9	49.6	2.9	513	5	PCT-US92-00282-20	Sequence 20, Appl
	10	46.2	2.7	51	5	PCT-US92-00282-37	Sequence 37, Appl
	11	44.2	2.6	1190	5	PCT-US92-00282-18	Sequence 18, Appl
	12	42.2	2.5	1669	2	US-08-522-421-1	Sequence 1, Appli
	13	40.6	2.4	1800	6	5180581-1	Patent No. 5180581
	14	40.6	2.4	2793	1	US-08-281-916-5	Sequence 5, Appli
	15	40.6	2.4	2793	2	US-08-460-725-7	Sequence 7, Appli
	16	38	2.2	1607	2	US-08-797-226-1	Sequence 1, Appli
	17	37.2	2.2	22846	2	US-08-469-461-3	Sequence 3, Appli
	18	37.2	2.2	22846	3	US-07-890-609-3	Sequence 3, Appli
	19	37	2.1	2781	3	US-08-749-522-4	Sequence 4, Appli
c	20	36.8	2.1	6152	4	US-08-973-462-1	Sequence 1, Appli
	21	36.6	2.1	1618	1	US-07-885-970A-24	Sequence 24, Appl
	22	36.6	2.1	1618	1	US-08-298-687A-24	Sequence 24, Appl
	23	36.6	2.1	1618	1	US-08-530-797-14	Sequence 14, Appl
	24	36.6	2.1	1618	1	US-08-298-829-24	Sequence 24, Appl
	25	36.6	2.1	1618	2	US-08-787-335-14	Sequence 14, Appl
	26	36.4	2.1	284	4	US-09-040-984-80	Sequence 80, Appl
	27	36	2.1	1219	5	PCT-US92-00282-8	Sequence 8, Appli
	28	35.6	2.1	5578	1	US-08-081-610-2	Sequence 2, Appli
	29	35	2.0	1627	3	US-09-106-464-1	Sequence 1, Appli
c	30	35	2.0	2749	1	US-08-344-695-1	Sequence 1, Appli
c	31	34.8	2.0	1075	4	US-08-960-780-30	Sequence 30, Appl
c	32	34.8	2.0	3095	6	5231168-1	Patent No. 5231168
	33	34.8	2.0	3336	4	US-09-330-970-4	Sequence 4, Appli
	34	34.6	2.0	790	6	5194596-8	Patent No. 5194596
c	35	34.4	2.0	669	2	US-09-109-266-7	Sequence 7, Appli
c	36	34.2	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	37	34	2.0	1183	2	US-08-731-722-8	Sequence 8, Appli
	38	33.8	2.0	1785	1	US-08-592-696-3	Sequence 3, Appli
	39	33.8	2.0	1785	2	US-09-027-536-3	Sequence 3, Appli
	40	33.8	2.0	1785	3	US-09-028-148-3	Sequence 3, Appli
	41	33.8	2.0	2265	1	US-08-242-689-1	Sequence 1, Appli
c	42	33.6	2.0	334	2	US-09-032-684-8	Sequence 8, Appli
c	43	33.4	1.9	1772	3	US-08-793-666-8	Sequence 8, Appli
c	44	33.4	1.9	1772	3	US-08-793-666-11	Sequence 11, Appl
	45	33.4	1.9	3781	3	US-08-688-988-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT   1
PCT-US92-00282-2
; Sequence 2, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2339 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-2


  Query Match           18.8%;  Score 324.2;  DB 5;  Length 2339;
  Best Local Similarity   52.7%;  Pred. No. 1.5e-80;
  Matches  804;  Conservative    0;  Mismatches  703;  Indels   18;  Gaps    4;

Qy      40 ctgatacatctcagttgttactttagctctgggagttgtggaaaagtgctggtgtgggcc 99
           ||| | |  |||    ||  |     ||  |     ||||||| ||| ||||| | ||
Db      46 CTGCTGCTCCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCC 105
```

```
Qy    100 gcagaatacagccattggatgaatatgaagacaatcctgaaagagcttgttcagagaggt 159
Db    106 ACTGATGGCAGCCCCTGGCTCAGCATGCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGC 165

Qy    160 catgaggtgactgtactggcatcttcagcttccattctttttgatcccaatgatgcatcc 219
Db    166 CACCAGGCGGTGGTCCTCACCCCAGAGGTGAATATGC------ACATCAAAGAAGAGAAA 219

Qy    220 actcttaaatttgaagtttatcctacatctttaactaaaactgaatttgagaatatcatc 279
Db    220 TTTTTCACCCTGACAGCCTATGCTGTTCCATGGACCCAGAAGGAATTTGATCGCGTTACG 279

Qy    280 atgcaacaggttaagagatggtcagacattcgaaaagatagcttttggttatatttttca 339
Db    280 CTGGGCTACACTCAAGGGT------TCTTTGAAACAGAACATCTTCTGAAGAGATATTCT 333

Qy    340 caagaacaagaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgta 399
Db    334 AGAAGTATGGCAATTATGAACAATGTATCTTTGGCCCTTCATAGGTGTTGTGTGGAGCTA 393

Qy    400 gtttcaaataagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgca 459
Db    394 CTGCATAATGAGGCCCTGATCAGCACCCTGAATGCTACTTCCTTTGATGTGGTTTTAACA 453

Qy    460 gatgctgtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtac 519
Db    454 GACCCCGTTAACCTCTGCGGGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTT 513

Qy    520 agtctccgctttactcctggctacacaattgaaaggcacagtggaggactgattttccct 579
Db    514 TTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGGGCACACAGTGTCCAAAT---CCT 570

Qy    580 ccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagg 639
Db    571 TCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGG 630

Qy    640 gtaaaaaatatgatctatgtgctttatttttgacttttggttccaaatgtctgatatgaag 699
Db    631 GTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCT 690

Qy    700 aagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagacaatg 759
Db    691 TATGCAAGTC---TTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTC 747

Qy    760 ggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatccattc 819
Db    748 AGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCGATC 807

Qy    820 ttaccaaacgttgattttgttggaggattccactgcaaacctgccaaacccctacctaag 879
Db    808 ATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAG 867

Qy    880 gaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgggg 939
Db    868 GAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGA 927

Qy    940 tcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaag 999
Db    928 TCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAA 987

Qy   1000 atcccacaaaaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaat 1059
Db    988 ATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAAC 1047

Qy   1060 actcggctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagct 1119
Db   1048 ACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCC 1107

Qy   1120 tttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatg 1179
Db   1108 TTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATG 1167

Qy   1180 gtgggcattccattgtttttgatcaacctgataacattgctcacatgaaggccaaggga 1239
Db   1168 GTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGA 1227

Qy   1240 gcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaag 1299
Db   1228 GCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAA 1287

Qy   1300 acagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacat 1359
Db   1288 GCAGTCATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCTCTCCAGCCTTCACAAG 1347

Qy   1360 gatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccac 1419
Db   1348 GACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCAC 1407

Qy   1420 aaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactct 1479
Db   1408 AAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCC 1467

Qy   1480 ttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaag 1539
Db   1468 TTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAA 1527

Qy   1540 ttttgtctgttttgtttctggaagt 1564
Db   1528 TGTTGTGCTTATGGCTACCGGAAAT 1552
```

RESULT 2
PCT-US92-00282-1
; Sequence 1, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2336 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US92-00282-1

Query Match 17.8%; Score 305.8; DB 5; Length 2336;

```
  Best Local Similarity   57.5%;  Pred. No. 1.9e-75;
  Matches  569;  Conservative    0;  Mismatches  417;  Indels    3;  Gaps    1;

Qy    576 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 635
          |||.  ||||||  | ||  |    | | ||          |||||  |||||  |||  || |
Db    564 CCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACATGACCTTCCTGCA 623

Qy    636 gagggtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 695
          | ||||  ||  ||  |||  ||   ||   |||      |||||  ||  |  |   ||  | ||
Db    624 GCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGCGACGTGGTTTATTC 683

Qy    696 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 755
            | |       |   ||  |   |||  |  |    |||      ||  |    ||
Db    684 CCCGTATGCAACCC---TTGCCTCAGAATTCCTTCAGAGAGAGGTGACTGTCCAGGACCT 740

Qy    756 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 815
          | ||  |    ||     | |||||  |  |||    ||          | |  |||   ||
Db    741 ATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGATTACCCTAGGCC 800

Qy    816 attcttaccaaacgttgattttgttggaggattccactgcaaacctgccaaacccctacc 875
            || | ||  ||  | | |||||||||  ||| || |||||   |     || || ||| |
Db    801 CATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCAAAATCCACTATC 860

Qy    876 taaggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctct 935
            |||||  | ||  |  |   |  |    |||||||||  ||||  |||||||  || ||| |
Db    861 CCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTT 920

Qy    936 ggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgc 995
          |||  |||  ||  |      | ||  |||    | |  ||  |  |  ||||||    ||  | |
Db    921 GGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGG 980

Qy    996 caagatcccacaaaaggttctgtggagatttgacgggaataaaccagatgccttaggtct 1055
          |||  |||||  ||  |   ||  ||||||  | |      ||  |     ||||        | |
Db    981 CAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAA 1040

Qy   1056 caatactcggctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccag 1115
          |||  ||     ||    ||||||  |||||||  ||  ||  ||  ||  |||||  ||  |   |||  |
Db   1041 CAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCG 1100

Qy   1116 agcttttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccc 1175
            ||  |||||  ||  ||||  |||   ||  ||||   |  |||||     ||  |  |  ||||   |  ||
Db   1101 TGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCC 1160

Qy   1176 tatggtgggcattccattgttttttgatcaacctgataacattgctcacatgaaggccaa 1235
          ||||||    ||  ||  ||||||  ||||||     ||  ||         | ||||  ||  |  ||
Db   1161 CATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAA 1220

Qy   1236 gggagcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcact 1295
          ||||||  |   ||  |    ||  |   |      ||||  |    |  |||  |    |||||  ||
Db   1221 GGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCT 1280

Qy   1296 gaagacagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattca 1355
            ||  ||||  ||  |||||        ||  ||  |||||  ||  |  |     |  ||  ||   ||||
Db   1281 AAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCTCTCCAGCCTTCA 1340

Qy   1356 acatgatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgcc 1415
            | ||  |  ||  ||  ||||  |||||  |  ||  ||  ||||||  |  ||  |||||  |||
Db   1341 CAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAG 1400

Qy   1416 ccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtacca 1475
          |||||  ||  ||   ||||||  ||    ||||||||  ||||||||||||||  |||||||||
Db   1401 GCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCA 1460

Qy   1476 ctctttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcac 1535
            ||  |||||  ||||||||  ||  ||   |||||    |||    ||  |||    ||  ||||  |
Db   1461 TTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTT 1520

Qy   1536 aaagttttgtctgttttgtttctggaagt 1564
            ||  | ||||    |  |  |  |  ||||  |
Db   1521 TAAATGTTGTGCTTATGGCTACCGGAAAT 1549


RESULT   3
```

```
PCT-US92-00282-25
; Sequence 25, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1561 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-25


  Query Match             6.2%;  Score 106.6;  DB 5;  Length 1561;
  Best Local Similarity   67.3%;  Pred. No. 3.4e-20;
  Matches  165;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy   1321 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1380
          || ||  |||||  ||  |||     | ||  ||  ||||  | ||  |  ||  ||  |||| |||
Db    272 TACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTG 331

Qy   1381 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1440
          ||  |  ||  ||  ||||||  |  ||  |||||  |||    |||||  ||  ||    ||||||  ||
Db    332 GACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGC 391

Qy   1441 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
              ||||||||  ||||||||||||||  |||||||||  ||  |||||  ||||||||  ||  ||
Db    392 CCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTC 451

Qy   1501 ctggcc-tgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1559
            |||||  | ||     ||  |||   ||  ||||  |   ||  | ||||   |  |  |  |  |
Db    452 TTGGCCGTCTGGCTGGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCG 511

Qy   1560 gaagt 1564
          |||  |
Db    512 GAAAT 516


RESULT   4
PCT-US92-00282-22
; Sequence 22, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
```

```
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  783 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-22


  Query Match             4.7%;  Score 81.6;  DB 5;  Length 783;
  Best Local Similarity   61.1%;  Pred. No. 2.2e-13;
  Matches  132;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy   1098 aggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatgaggc 1157
          |||||| || |  ||| | || ||||| || |||| |||  || ||||  | |||||
Db    470 AGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAG 529

Qy   1158 aatctaccatgggatccctatggtgggcattccattgtttttttgatcaacctgataacat 1217
           || | | ||||  | || ||||||   || || ||||||  ||||||    || ||
Db    530 CATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGC 589

Qy   1218 tgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagtac 1277
             | |||| || | |||||||| |  || |  || |  |      |||| |   |
Db    590 AAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGA 649

Qy   1278 agacctgctgaatgcactgaagacagtaattaatga 1313
          |||  |    ||||| || ||  |||| || |||||
Db    650 AGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGA 685


RESULT   5
PCT-US92-00282-10
; Sequence 10, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-10


  Query Match             3.7%;  Score 64.2;  DB 5;  Length 1008;
  Best Local Similarity   45.8%;  Pred. No. 1.7e-08;
  Matches  385;  Conservative    0;  Mismatches  438;  Indels   18;  Gaps    4;

Qy     40 ctgatacatctcagttgttactttagctctgggagttgtggaaaagtgctggtgtgggcc 99
          ||| | | ||||    ||  |     ||  ||  |    |||| || |||||||| | ||
Db    157 CTGCTGCTTCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGGAAGGTGCTGGTGGTGCCC 216

Qy    100 gcagaatacagccattggatgaatatgaagacaatcctgaaagagcttgttcagagaggt 159
           | ||   |||||| ||| | |  |||  |     | ||   || ||   |  ||||||
Db    217 ACTGATGGCAGCCACTGGCTCAGCATGCGGGAGGCCTTGCGGGACCTCCATGCGAGAGGC 276

Qy    160 catgaggtgactgtactggcatcttcagcttccattctttttgatcccaatgatgcatcc 219
          ||  |||||   || ||  |       |     |   | | |   ||| || |    |
Db    277 CACCAGGTGGTGGTCCTCACCCTGGAGG------TGAATATGTACATCAAAGAAGAGAAC 330

Qy    220 actcttaaatttgaagtttatcctacatctttaactaaaactgaatttgagaatatcatc 279
           | | |   |   |   ||| | |  || |  ||  |    ||||||||         |
Db    331 TTTTTCACCCTGACAACGTATGCCATTTCATGGACCCAGGACGAATTTGA------TCGC 384

Qy    280 atgcaacaggttaagagatggtcagacattcgaaaagatagcttttggttatatttttca 339
           |    | || | | |     ||   | ||  || |||     ||  | | | |||||
Db    385 CTTTTGCTGGGTCACACTCAATCGTTCTTTGAAACAGAACATCTTCTGATGAAATTTTCT 444

Qy    340 caagaacaagaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgta 399
           |  |   | |||  ||    |  | | |   ||  |    |  |  |||   ||  ||
Db    445 AGAAGAATGGCAATTATGAACAATATGTCTTTGATCATACATAGGTCTTGTGTGGAGCTA 504

Qy    400 gtttcaaataagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgca 459
           |    ||| ||    | || | | || || |   |    |||||  | ||| |  ||
Db    505 CTGCATAATGAGGCCCTGATCAGGCACCTGCATGCTACTTCCTTTGATGTGGTTCTAACA 564

Qy    460 gatgctgtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtac 519
          ||  |  ||   | ||| |  | | |||||||| |   ||    || |   |||||
Db    565 GACCCCTTTCACCTCTGCGCGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTT 624

Qy    520 agtctccgctttactcctggctacacaattgaaaggcacagtggaggactgatttttccct 579
             |  |    | |||  |  |   |       | |  ||       | |  |    ||||
Db    625 TTCTTGAGGAACATTCCATGTGATTTAGACTTTAAGGGCACACAGTGTCCAA---ACCCT 681

Qy    580 ccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagg 639
           | ||||| || ||||   |  |  | |        || || ||||| ||| || | |||
Db    682 TCCTCCTATATTCCTAGATTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGG 741

Qy    640 gtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatatgaag 699
```

```
          || || || ||| ||||    ||     |   ||     | |||     ||| |
Db    742 GTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACCTTTGCCA---TGCTGTTTCTGCT 798

Qy    700 aagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagacaatg 759
          | |       ||  |  |||  ||||    |||        |  |  | ||     |
Db    799 CCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTC 858

Qy    760 ggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatccattc 819
          |  | ||   |  |||||  | |||       |       |  | |  ||    ||  ||
Db    859 AGCCATGCATCCTGGTGGCTGTTCCGAGGGGAGTTTGTGATGGATTACCCCAGGCCGATC 918

Qy    820 ttaccaaacgttgattttgttggaggattccactgcaaacctgccaaacccctacctaag 879
          | || ||| | |  ||  |||| ||  || ||||       |  || || ||| || ||
Db    919 ATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAG 978

Qy    880 g 880
          |
Db    979 G 979


RESULT   6
PCT-US92-00282-14
; Sequence 14, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-14


  Query Match             3.4%;  Score 59.4;  DB 5;  Length 1066;
  Best Local Similarity   47.2%;  Pred. No. 3.7e-07;
  Matches  291;  Conservative    0;  Mismatches  311;  Indels   15;  Gaps    3;

Qy     40 ctgatacatctcagttgttactttagctctgggagttgtggaaaagtgctggtgtgggcc 99
          ||| | | ||||    ||  |     ||  |      ||||||| ||| |||||  | ||
Db    176 CTGCTGCTTCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCC 235

Qy    100 gcagaatacagccattggatgaatatgaagacaatcctgaaagagcttgttcagagaggt 159
                |  |||||| ||| | |  |||  |    || ||   |||||   |   |||||

Db    236 ATTCATGGCAGCCACTGGCTCAGCATGCGGGAGGTCTTGCGGGAGCTCCATGCCAGAGGC 295

Qy    160 catgaggtgactgtactggcatcttcagcttccattctttttgatcccaatgatgcatcc 219
          ||  |||     || ||  |  |      |    || |     |   ||| || |  | |
Db    296 CACCAGGCAGTGGTCCTCACCCCAGAGGTGAATATGC------ACATCAAAGAAGAGAAC 349

Qy    220 actcttaaatttgaagtttatcctacatctttaactaaaactgaatttgagaatatcatc 279
          | | |   |   |   ||| | |  || |  ||  |  |||||||||         |
Db    350 TTTTTCACCCTGACAACCTATGCCATTTCGTGGACCCAGGATGAATTTGATCGCCATGTG 409

Qy    280 atgcaacaggttaagagatggtcagacattcgaaaagatagcttttggttatatttttca 339
          ||   ||   | |   |       || ||  || |||    |||  |    | ||||
Db    410 CTGGGCCACACTCAACTGT------ACTTTGAAACAGAACATTTTCTGAAGAAATTTTTC 463

Qy    340 caagaacaagaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgta 399
          |     | |||  ||    |  | | |    | | |    |  |  |||   ||  ||
Db    464 AGAAGTATGGCAATGTTGAACAATATGTCTTTGGTCTATCATAGGTCTTGTGTGGAGCTA 523

Qy    400 gtttcaaataagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgca 459
          |    ||| ||    | || |   | ||  | |   |    |||||  | |||||  ||
Db    524 CTACATAATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACA 583

Qy    460 gatgctgtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtac 519
          ||  | |||   | ||| |  |   |||||||| |   ||    || |    |||||
Db    584 GACCCCGTTAACCTCTGCGCGGCAGTGCTGGCTAAGTACCTGTCGATTCCTACTGTGTTT 643

Qy    520 agtctccgctttactcctggctacacaattgaaaggcacagtggaggactgattttccct 579
          | |  |    | |||  |  |   |      | |  ||      | |  |    ||||
Db    644 TTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGGGCACACAGTGTCCAA---ACCCT 700

Qy    580 ccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagg 639
          | ||||| || ||||   |  |  |||         || || ||||| |||||| | |||
Db    701 TCCTCCTATATTCCTAGATTACTAACAACCAATTCAGACCACATGACATTCATGCAAAGG 760

Qy    640 gtaaaaaatatgatcta 656
          || || || ||| ||||
Db    761 GTCAAGAACATGCTCTA 777


RESULT   7
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1197 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-12


  Query Match             3.4%;  Score 58;  DB 5;  Length 1197;
  Best Local Similarity   47.0%;  Pred. No. 9.5e-07;
  Matches  250;  Conservative    0;  Mismatches  275;  Indels    7;  Gaps    2;

Qy    349 gaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaat 408
Db    573 GCAATTATGAACAATGTATCTTTGGCCCTTCATAGGTGTTGTGTGGAGCTACTGCATAAT 632

Qy    409 aagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtt 468
Db    633 GAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCCGTT 692

Qy    469 tttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgc 528
Db    693 AACCTCTGTGGGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTTTTTTGGAGG 752

Qy    529 tttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctac 588
Db    753 TACATTCCATGTGACTTAGACTTTAAGGCCATACAGTGTCAAAT----CCTTCCTCCTAT 808

Qy    589 atacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaat 648
Db    809 ATTCCTAAGTTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAAC 868

Qy    649 atgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggat 708
Db    869 ATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCTTATGCAAGT 928

Qy    709 cagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagct 768
Db    929 C---TTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTCAGCTATGCA 985

Qy    769 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 828
Db    986 TCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCGATCATGCCCAAC 1045

Qy    829 gttgattttgttggaggattccactgcaaacctgccaaaccccctacctaagg 880
Db   1046 ATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAGG 1097


RESULT   8
PCT-US92-00282-16
; Sequence 16, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1448 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
PCT-US92-00282-16


  Query Match             3.1%;  Score 52.8;  DB 5;  Length 1448;
  Best Local Similarity   45.7%;  Pred. No. 2.9e-05;
  Matches  263;  Conservative    0;  Mismatches  307;  Indels    6;  Gaps    2;

Qy    305 acattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgggaat 364
Db    404 ACTTTGAAACAGAACATTTTCTGAAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATT 463

Qy    365 tatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaa 424
Db    464 CATCTTTGGTCTTTCATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGC 523

Qy    425 aactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggtgagc 484
Db    524 ACCTGAATGCTACTTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCTGTGGGGCAG 583

Qy    485 tgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggctaca 544
Db    584 TGCTGGCTAAGTACCTATCAATTCCTGCTGTGTTTTTTTTGAGGAACATTCCATGTGATT 643

Qy    545 caattgaaaggcacagtggaggactgattttccctccttcctacatacctattgttatgt 604
Db    644 TAGACTTTAAGGGCACACAGTGTCCAAAT---CCTTCCTCCTATATTCCTAAGTTACTAA 700

Qy    605 caaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtgcttt 664
Db    701 CGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGG 760

Qy    665 attttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagtgaag 724
Db    761 CCCTGTCCTACCTTTGCCA---TGCTCTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGC 817

Qy    725 ttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgc 784
Db    818 TTTTTCAGAGGGAGGTGTCAGTGGTGGATCTTCTCAGCCATGCATCTGTGTGGCTGTTCC 877

Qy    785 gaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggag 844
Db    878 GAGGGGACTTTGTGATGGATTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGG 937

Qy    845 gattccactgcaaacctgccaaaccccctacctaagg 880
Db    938 GCATCAACTGTGCCAACAGGAAGCCACTATCTCAGG 973


RESULT   9
PCT-US92-00282-20
; Sequence 20, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
```

```
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-20


  Query Match             2.9%;  Score 49.6;  DB 5;  Length 513;
  Best Local Similarity   60.3%;  Pred. No. 0.00014;
  Matches   82;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     878 aggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgg 937
           ||||| | || |  |   | |    ||||||||| ||||  ||||||| || ||| |||
Db     229 AGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGG 288

Qy     938 ggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgcca 997
           | ||| || |     | ||  |||   | |  || | | ||||||    ||  | | ||
Db     289 GATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCA 348

Qy     998 agatcccacaaaaggt 1013
           | ||||| || |  ||
Db     349 AAATCCCTCAGACAGT 364


RESULT  10
PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-37


  Query Match             2.7%;  Score 46.2;  DB 5;  Length 51;
  Best Local Similarity   94.1%;  Pred. No. 0.0004;
  Matches   48;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1450 catgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1500
           || |||||||||||||||||||||||||||||||||||||||||||||||| || |||
Db       1 CACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGTTTCCTG 51


RESULT  11
PCT-US92-00282-18
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18
```

```
  Query Match             2.6%;  Score 44.2;  DB 5;  Length 1190;
  Best Local Similarity   48.9%;  Pred. No. 0.0065;
  Matches  149;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy     576 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 635
           |||    |||||| | || |    | | ||          ||||| ||||| ||| || |
Db     648 CCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACATGACCTTCCTGCA 707

Qy     636 gagggtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 695
           | |||| || || ||| ||  ||   |||      |||||  || | |  ||  | ||
Db     708 GCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGCGACGTGGTTTATTC 767

Qy     696 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 755
           | |       |   ||  |  ||| |  |    |||       ||  |    ||
Db     768 CCCGTATGCAACCC---TTGCCTCAGAATTCCTTCAGAGAGAGGTGACTGTCCAGGACCT 824

Qy     756 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 815
           | || |   ||     | ||||  |  |||    ||         |    |||   ||
Db     825 ATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGATGACCCTAGGCC 884

Qy     816 attcttaccaaacgttgattttgttggaggattccactgcaaacctgccaaacccctacc 875
           || | || ||  | | |||||||||| ||| || ||||   |.    || || ||| |
Db     885 CATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCAAAATCCACTATC 944

Qy     876 taagg 880
           |||
Db     945 CCAGG 949


RESULT  12
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
;  GENERAL INFORMATION:
;    APPLICANT:  Abu-Bakar, Umi Kalsom
;    APPLICANT:  Barton, Sarah Louise
;    APPLICANT:  Gallego-Veigas, Pedro Pablo
;    APPLICANT:  Gray, Julie Elizabeth
;    APPLICANT:  Grierson, Donald
;    APPLICANT:  Lowe, Alexandra Louise
;    APPLICANT:  Picton, Steve
;    APPLICANT:  Whotton, Lee Colin
;    TITLE OF INVENTION:  DNA, DNA CONSTRUCTS, CELLS AND PLANTS
;    TITLE OF INVENTION:  DERIVED THEREFROM
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/522,421
;      FILING DATE:  11-JAN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305868.3
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305869.1
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305859.2
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305865.9
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305866.7
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305867.5
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305860.0
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305862.6
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9314351.9
;      FILING DATE:  12-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9320988.0
;      FILING DATE:  12-OCT-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  ERT1B
US-08-522-421-1


  Query Match             2.5%;  Score 42.2;  DB 2;  Length 1669;
  Best Local Similarity   49.1%;  Pred. No. 0.027;
  Matches  141;  Conservative    0;  Mismatches  143;  Indels    3;  Gaps    1;

Qy    1072 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1131
           || ||| ||| || ||  |  | || |  |||||          ||  ||| | ||||||
Db     972 AAATGGGCACCACAAAAACAGGTGCTTGCACATCCCGCGGTAGCAGGGTTTTTCACTCAT 1031

Qy    1132 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1191
            ||||    |||   |    |||    || |   | |   |||||||||||| |||  |||
Db    1032 TGTGGTTGGAATTCTACGCTTGAAAGTATATGTGAAGAAGTCCCTATGGTGTGCAGGCCA 1091

Qy    1192 ttgttttttgatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1251
           ||  |    || ||||  |  |||      |  |||   | |  |       ||| |
Db    1092 TTTCTAGCAGACCAACTGGTGAACGCAAGGTATCTGAG---CCAAATATACAAGGTTGGG 1148

Qy    1252 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1311
           || ||||  |    ||   ||| || | | |   |||  || | |  |  | || ||
Db    1149 TTCGAATTGGAGGTTATCGAGAGAACGGTCATAGAGAAAACAATAAGAAAACTCATGTTA 1208

Qy    1312 gatcctttatataaagagaatattatgaaattatcaagaattcaaca 1358
              |     |  |||||    |  |  | | || ||   ||  ||||
Db    1209 AGTGAAGAAGGCAAAGATGTGAAGAAAAGAGTAGCAGACATGAAACA 1255


RESULT  13
5180581-1
;Patent No. 5180581
;    APPLICANT: MILLER, LOIS K.;O'REILLY, DAVID R.
;    TITLE OF INVENTION: BIOLOGICAL INSECT CONTROL AGENTS AND
;METHODS OF USE
;    NUMBER OF SEQUENCES: 2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/373,952
;      FILING DATE: 29-JUN-1989
;SEQ ID NO:1:
;      LENGTH: 1800
5180581-1


  Query Match             2.4%;  Score 40.6;  DB 6;  Length 1800;
```

```
  Best Local Similarity   48.5%;  Pred. No. 0.079;
  Matches  112;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy   1068 gtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1127
          | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
Db   1186 gcaaaattggtttaatcaacgcgccgtgctgcgtcataaaaaaatggcggcgtttattac 1245

Qy   1128 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1187
           || || |||    |    | | | |||||  |  |    |||||  || |||||| |  |
Db   1246 gcaaggcggactacaatcgagcgacgaggccttggaagccgggataccatggtgtgtct 1305

Qy   1188 tccattgttttttgatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1247
           ||  || |    || ||   | |  |        |  |  ||      || | |||
Db   1306 gcccatgatgggcgaccagttttaccatgcgcacaaattacagcaactcggcgtagcccg 1365

Qy   1248 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1298
             ||||||      ||  | || ||     | || ||    ||  | ||
Db   1366 cgccttggacactgttaccgtttccagcgatcaactactagtggcgataaa 1416


RESULT  14
US-08-281-916-5
; Sequence 5, Application US/08281916
; Patent No. 5662897
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Lois K.
;    APPLICANT:  Black, Bruce C.
;    APPLICANT:  Dierks, Peter M.
;    APPLICANT:  Fleming, Nancy C.
;    APPLICANT:  Ahmed, Fakhruddin
;    TITLE OF INVENTION:  Insect Viruses, Sequences, Insecticidal
;    TITLE OF INVENTION:  Compositions and Methods of Use
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/281,916
;      FILING DATE:  27-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  29-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
US-08-281-916-5


  Query Match             2.4%;  Score 40.6;  DB 1;  Length 2793;
  Best Local Similarity   48.5%;  Pred. No. 0.098;
  Matches  112;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy   1068 gtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1127
```

```
          | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
Db   1186 GCAAAATTGGTTTAATCAACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTATTAC 1245

Qy   1128 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1187
           || || |||    |    | | | |||||  |  |    |||||  || |||||| |  |
Db   1246 GCAAGGCGGACTACAATCGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCT 1305

Qy   1188 tccattgttttttgatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1247
           ||  || |    || ||   | |  |        |  |  ||      || | |||
Db   1306 GCCCATGATGGGCGACCAGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCG 1365

Qy   1248 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1298
             ||||||      ||  | || ||     | || ||    ||  | ||
Db   1366 CGCCTTGGACACTGTTACCGTTTCCAGCGATCAACTACTAGTGGCGATAAA 1416


RESULT  15
US-08-460-725-7
; Sequence 7, Application US/08460725
; Patent No. 5858353
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Lois K.
;    APPLICANT:  Black, Bruce C.
;    APPLICANT:  Dierks, Peter M.
;    APPLICANT:  Fleming, Nancy C.
;    APPLICANT:  Ahmed, Fakhruddin
;    TITLE OF INVENTION:  Insect Viruses, Sequences, Insecticidal
;    TITLE OF INVENTION:  Compositions and Methods
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,725
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/281,916
;      FILING DATE:  27-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  29-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
US-08-460-725-7


  Query Match             2.4%;  Score 40.6;  DB 2;  Length 2793;
  Best Local Similarity   48.5%;  Pred. No. 0.098;
  Matches  112;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy   1068 gtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1127
          | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
```

```
Db   1186 GCAAAATTGGTTTAATCAACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTATTAC 1245

Qy   1128 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1187
          || || |||    |    | | | |||||  |  |    ||||| || |||||| |  |
Db   1246 GCAAGGCGGACTACAATCGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCT 1305

Qy   1188 tccattgttttttgatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1247
          ||  || |    || ||   | |  |        |  |  ||      || | |||
Db   1306 GCCCATGATGGGCGACCAGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCG 1365

Qy   1248 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1298
             ||||||      ||  | || ||     | || ||    ||  | ||
Db   1366 CGCCTTGGACACTGTTACCGTTTCCAGCGATCAACTACTAGTGGCGATAAA 1416
```

Search completed: May 1, 2001, 00:34:00
Job time: 56584 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:17:51 ; Search time 12176.5 Seconds
(without alignments)
20.670 Million cell updates/sec

Title: US-09-721-183-18
Perfect score: 1722
Sequence: 1 tgcaccaggatgactctgaa.....acaaaaaaaaaaaaaaaaaa 1722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	684	39.7	823	151	BF689099	BF689099 602185172
	2	580	33.7	746	151	BF688309	BF688309 602185172
	3	541.4	31.4	1005	174	BG167171	BG167171 602344696
	4	456.2	26.5	513	144	BF105272	BF105272 601822259
c	5	438	25.4	761	16	AI118428	AI118428 ue40e09.x
	6	427.4	24.8	651	4	AA255327	AA255327 mz84f10.r
c	7	424.6	24.7	669	102	AI821970	AI821970 ob22c12.x
	8	423.6	24.6	923	21	AI529976	AI529976 ui87g04.y
	9	421.4	24.5	1192	169	BF789664	BF789664 602105279
c	10	410.4	23.8	482	14	AA987868	AA987868 oq81f04.s
	11	409.8	23.8	996	21	AI530408	AI530408 ui92f08.y
	12	408.2	23.7	616	4	AA255225	AA255225 mz85d04.r
c	13	397.6	23.1	517	117	AW604423	AW604423 RC3-CT025
	14	396	23.0	792	169	BF786091	BF786091 602110662
	15	390.4	22.7	628	31	AV657530	AV657530 AV657530
	16	387	22.5	975	149	BF533975	BF533975 602075137
	17	383.6	22.3	924	21	AI528431	AI528431 ui97d02.y
	18	383.2	22.3	410	31	AV696157	AV696157 AV696157

	19	381	22.1	554	161	BE013295 BE013295 123199 MA
	20	374.2	21.7	819	16	AI156939 AI156939 ui49c01.y
	21	373.8	21.7	909	21	AI528459 AI528459 ui97f08.y
	22	373.6	21.7	934	169	BF786200 BF786200 602110490
c	23	372.2	21.6	432	14	AA962194 AA962194 oq69h06.s
	24	371	21.5	762	11	AA717701 AA717701 vp94g02.r
	25	369.8	21.5	781	21	AI529761 AI529761 ui82b06.y
c	26	368.8	21.4	611	204	AQ315601 AQ315601 RPCI11-97
c	27	359.8	20.9	639	146	BF322562 BF322562 maa28e12.
	28	359.6	20.9	796	16	AI097842 AI097842 ue40e09.y
c	29	342.8	19.9	440	111	AW129268 AW129268 xf22g02.x
c	30	342	19.9	601	205	AQ376853 AQ376853 RPCI-11-1
c	31	340.6	19.8	425	165	BE326864 BE326864 hr66b05.x
	32	340.6	19.8	1096	169	BF788115 BF788115 602113339
c	33	340.4	19.8	565	11	AA746229 AA746229 ob22c12.s
	34	334.4	19.4	453	32	AV719818 AV719818 AV719818
	35	331.4	19.2	557	143	BF080461 BF080461 231360 MA
c	36	330.6	19.2	589	18	AI256545 AI256545 ui87g04.x
	37	329.8	19.2	966	143	BF027320 BF027320 601671692
c	38	329.4	19.1	643	102	AI785430 AI785430 uj42b06.x
	39	328	19.0	456	227	B86874 B86874 RPCI11-26H1
	40	327.6	19.0	777	16	AI116747 AI116747 ue28e05.y
	41	326.6	19.0	413	31	AV655524 AV655524
c	42	326.4	19.0	836	21	AI529553 AI529553 ui82a02.x
c	43	324.6	18.9	530	21	AI480418 AI480418 tm52b06.x
c	44	324.2	18.8	555	18	AI265736 AI265736 uj05f12.x
	45	322.6	18.7	844	22	AI573962 AI573962 uj66f05.y

ALIGNMENTS

```
RESULT   1
BF689099/c
LOCUS       BF689099      823 bp    mRNA            EST       22-DEC-2000
DEFINITION  602185172T1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299880 3',
            mRNA sequence.
ACCESSION   BF689099
VERSION     BF689099.1  GI:11974507
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1157  row: o  column: 17
            High quality sequence start: 5
            High quality sequence stop: 796.
FEATURES             Location/Qualifiers
     source          1. .823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4299880"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      221 a    199 c    164 g    239 t
ORIGIN


  Query Match             39.7%;  Score 684;  DB 151;  Length 823;
  Best Local Similarity   92.3%;  Pred. No. 1.1e-163;
  Matches  742;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    2;

Qy     854 gcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacagagctctggagaaaa 913
           ||||||| ||||| ||| |  | | ||||||||| || |||||||||| ||||||||||
Db     804 GCAAACCGGCCAACCCCTGCCTAAGGGAAATGGAAGACTTTGTACAGAGTTCTGGAGAAA 745

Qy     914 atggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagggcca 973
           |||||||||||||||||||||||||||||| || | |||||||||||| ||||||||||
Db     744 ATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAAAGGGCCA 685

Qy     974 atgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgacggga 1033
           | ||||||||| ||| ||| ||| ||||||||||||||||||||||||||||||| ||||
Db     684 ACGTAATTGCATCAGACCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTGATGGGA 625

Qy    1034 ataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccagaatgacc 1093
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     624 ATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACCCCAGAATGACC 565

Qy    1094 ttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatg 1153
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||| |
Db     564 TTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGCATCTACG 505

Qy    1154 aggcaatctaccatgggatccctatggtgggcattccattgtttttgatcaacctgata 1213
           |||||||||||||||||||||||||||||||| ||||||||||||   ||||||||||||
Db     504 AGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCAACCTGATA 445

Qy    1214 acattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcga 1273
           |||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db     444 ACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGGACTTCAACACAATGTCGA 385

Qy    1274 gtacagacctgctgaatgcactgaagacagtaattaatgatcctttatataaagagaata 1333
           |||||||| |||||||||| |||||| |||||||||||||||| ||||||||||||
Db     384 GTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATCCTTCATATAAAGAGAATG 325

Qy    1334 ttatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgagcagtct 1393
           |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||
Db     324 TTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAG-CCCTGGATCGAGCAGTCT 266

Qy    1394 tctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcagcccatg 1453
           |||||||||||||||||||| |||||||||| |||||||||| |||||||||||| |
Db     265 TCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGGGTTGCAGCCCACG 206

Qy    1454 acctcacctggttccagtaccactctttggatgtgattgggtttctgctggcctgtgtgg 1513
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Db     205 ACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTGCTGGTCTGTGTGG 146

Qy    1514 caactgtgatatttatcatcacaaagttttgtctgttttgtttctggaagtttgctagaa 1573
           |||||||||||||||| ||||||| | ||||||||||||||||||||||||||||||||
Db     145 CAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTTTCTGGAAGTTTGCTAGAA 86

Qy    1574 aagggaagaagggaaaaagagattagttatgtctgacatttgaagctggaaaaccaga-t 1632
           |||  |||||||||||||  ||||||||||  ||||| |||||||||||||||||||| || |
Db      85 AAGCAAAGAAGGGAAAAAATGATTAGTTATATCTGAGATTTGAAGCTGGAAAACCTGACT 26

Qy    1633 agataggacaacttcagtttattc 1656
           || |  ||| |||||||||  | |
Db      25 AGGTGAGACTACTTCAGTTATTCC 2


RESULT   2
BF688309
LOCUS       BF688309      746 bp    mRNA            EST       22-DEC-2000
DEFINITION  602185172F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299880 5',
```

```
            mRNA sequence.
ACCESSION   BF688309
VERSION     BF688309.1  GI:11973717
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1157  row: o  column: 17
            High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4299880"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      212 a    158 c    183 g    193 t
ORIGIN


  Query Match           33.7%;  Score 580;  DB 151;  Length 746;
  Best Local Similarity   92.4%;  Pred. No. 3.4e-137;
  Matches  665;  Conservative    0;  Mismatches   50;  Indels    5;  Gaps    5;

Qy    835 tttgttggaggattccactgcaaacctgccaaaccctacctaaggaaatggaggagttt 894
          |||||||||||| |||||||||||||||||||||||| |||||||||||||| || |||
Db      3 TTTGTTGGAGGACTCCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTT 62

Qy    895 gtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaac 954
          |||||||||||||||||||||||||||||||||||||||||||||||| || | ||||||
Db     63 GTACAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAAC 122

Qy    955 atgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggtt 1014
          ||||||| |||||||||||| |||||||| ||||||| ||| ||||||||||||||||
Db    123 ATGACAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTT 182

Qy   1015 ctgtggagatttgacgggaataaaccagatgccttaggtctcaatactcggctgtacaag 1074
          ||||||||||||| |||||||||||||| |||||||||||||||||||||||||| |||
Db    183 CTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAG 242

Qy   1075 tggataccccagaatgaccttctaggtcatccaaaaaccagagctttttataactcatggt 1134
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    243 TGGATACCCCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGT 302

Qy   1135 ggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattg 1194
          |||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db    303 GGAGCCAATGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTG 362

Qy   1195 ttttttgatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattg 1254
```

```
          |||   |||||||||||||||||||||||||||||||| |||||||||||||||| ||
Db    363 TTTGCCGATCAACCTGATAACATTGCTCACATGAAGGCC-AGGGAGCAGCTGTTAGAGTG 421

Qy   1255 gacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgat 1314
          |||||||||||||||||||||||||||||| |||||||||| |||||| ||||||||||||
Db    422 GACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGAT 481

Qy   1315 cctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaag 1374
          |||| |||||||||||||| |||||||||||||||||||||||||||||||||||| |||
Db    482 CCTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAG 541

Qy   1375 cccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacac 1434
          |||||||||||||||||||||||| |||||||||||| |||||||||||| ||||||
Db    542 CCCCTGGATCGAGCAGTCTTCTGGATCGAATTTGTCATGCGCCACAAAGGAGCTAAACAC 601

Qy   1435 cttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattggg 1494
          ||||| |||||||||||| || |||||||||||||||||||||| ||||||||||||||
Db    602 CTTCGGGTTGCAGCCCACGA-CTCACCTGGTTCCAGTACCACTC-TTGGATGTGATTGGG 659

Qy   1495 tttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgt 1554
          |||||||| |||||||| ||||||||||         || ||| | | | | || |
Db    660 -TTCTGCTGGTCTGTGTGGC-ACTGTGATATTATCGTCACAAATGTTGTCTGTCTGTTCT 717


RESULT   3
BG167171
LOCUS       BG167171     1005 bp    mRNA            EST       06-FEB-2001
DEFINITION  602344696F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455022 5',
            mRNA sequence.
ACCESSION   BG167171
VERSION     BG167171.1  GI:12673874
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247  row: o  column: 23
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4455022"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      256 a    252 c    252 g    244 t      1 others
ORIGIN


  Query Match           31.4%;  Score 541.4;  DB 174;  Length 1005;
  Best Local Similarity   85.0%;  Pred. No. 2.6e-127;
  Matches  653;  Conservative    0;  Mismatches  107;  Indels    8;  Gaps    4;
```

```
Qy    727 ttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcga 786
          || ||||||||||||||| |||| |||||||||||| |||||||| |||||||||| |||
Db    103 TTTGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTATTCGA 162

Qy    787 aactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggagga 846
          |||||||||| |||||| |||||||||||| |||||||||| |||||||||||||||||
Db    163 AACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATTTTGTTGGAGGA 222

Qy    847 ttccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctct 906
          ||||||||||||||||||||||||| |||||||||||||| || ||||||||||||||||
Db    223 CTCCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAGCTCT 282

Qy    907 ggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaa 966
          |||||||||||||||||||||||||||||||||||||| || | |||||||||||| ||||
Db    283 GGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAA 342

Qy    967 agggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagattt 1026
          |||||||| |||||||| ||||||| ||| ||||||||||||||||||||||||||||||
Db    343 AGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATTT 402

Qy   1027 gacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccag 1086
          || |||||||||||||| ||||||||||||||||||||||| ||||||||||||||||
Db    403 GATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACCCCAG 462

Qy   1087 aatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggc 1146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 AATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGC 522

Qy   1147 atctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttgatcaa 1206
          ||||| ||||||||||||||||||||||||||||||| ||||||||||||   |||||
Db    523 ATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCAN 582

Qy   1207 cctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc-aacac 1265
          ||||||||||||||||||||||||||| | |||||||||||||| |||||||| |||||
Db    583 ACTGATAACATTGCTCACATGAAGGCCAGGCGAGCAGCTGTTAGAGTGGACTTCAAACAC 642

Qy   1266 aatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttttatataa 1325
          |||||||||||||| |||||||||  ||| || ||||||| |||||||| |||| |
Db    643 AATGTCGAGTACAGACTTGCTGAATGC-TTGACGAGAGTAATTCATGATCCTTCATATCA 701

Qy   1326 agagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcg 1385
          |||||| |    | | | ||   | |      | ||||        |  ||| ||||||
Db    702 CGAGAATGTCTGCACACTCTCCGAATTCACTGGTCATCA-----GTGAAGCCCCGGATCG 756

Qy   1386 agcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgc 1445
          ||||| |||| |||||||| ||| |||||  |  || || |  ||||   | |   ||||
Db    757 AGCAGGCTTCCGGATTGAA-TTGGCATGCGGCCAAAGGGCGTAAAACCTTTGCTGCTTGC 815

Qy   1446 agcccatgacctcacctggttccagtaccactctttggatgtgattgg 1493
            ||||  | |    | |||||| || || |||| |    |||| ||
Db    816 GTCCCAGAAACCTCACCGGTTCCCGTCCCCCTCTGGGAGGTTGATGGG 863


RESULT   4
BF105272
LOCUS       BF105272      513 bp    mRNA            EST       19-OCT-2000
DEFINITION  601822259F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4042383 5',
            mRNA sequence.
ACCESSION   BF105272
VERSION     BF105272.1  GI:10887798
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM863  row: f  column: 16
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4042383"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      147 a    116 c    121 g    129 t
ORIGIN


  Query Match             26.5%;  Score 456.2;  DB 144;  Length 513;
  Best Local Similarity   93.5%;  Pred. No. 1.2e-105;
  Matches  476;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy    729 aggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaa 788
          |||||||||||||||| |||| |||||||||||| ||||||||| |||||||||| |||||
Db      5 AGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTATTCGAAA 64

Qy    789 ctcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggatt 848
          |||||||| |||||| |||||  ||||| |||||||||| |||||||||||||||||| |
Db     65 CTCCTGGAATTTTCAGTTTCCATATCCACTCTTACCAAATGTTGATTTTGTTGGAGGACT 124

Qy    849 ccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctctgg 908
          |||||||||||||||||||||||| |||||||||||||| || ||||||||||||||||
Db    125 CCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAGCTCTGG 184

Qy    909 agaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaag 968
          |||||||||||||||||||||||||||||||||||||| || | |||||||||||| ||||||
Db    185 AGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAAAG 244

Qy    969 ggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttga 1028
          |||||| |||||||| ||||||| ||| ||||||||||||||||||||||||||||||||
Db    245 GGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTGA 304

Qy   1029 cgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccagaa 1088
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db    305 TGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTCTACAAGTGGATACCCCAGAA 364

Qy   1089 tgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcat 1148
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    365 TGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGCAT 424

Qy   1149 ctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttgatcaacc 1208
          ||| ||||||||||||||||||||||||||||||| ||||||||||||   ||||||||
Db    425 CTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCAACC 484

Qy   1209 tgataacattgctcacatgaaggccaagg 1237
          |||||||||||||||||||||||||||| ||
Db    485 TGATAACATTGCTCACATGAAGGCCAGGG 513


RESULT   5
AI118428/c
LOCUS       AI118428      761 bp    mRNA            EST       02-SEP-1998
```

```
DEFINITION  ue40e09.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1482856 3' similar to gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AI118428
VERSION     AI118428.1  GI:3518752
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:931212
            Seq primer: custom primer used
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .761
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1482856"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      198 a    168 c    165 g    230 t
ORIGIN
```

```
  Query Match             25.4%;  Score 438;  DB 16;  Length 761;
  Best Local Similarity   74.7%;  Pred. No. 5.6e-101;
  Matches  563;  Conservative    0;  Mismatches  190;  Indels    1;  Gaps    1;

Qy     865 aaacccctacctaagga-aatggaggagtttgtacagagctctggagaaaatggtgttgt 923
           |||||| | |||||||| |||| | |||| |||||||||| ||| |||||| ||
Db     761 AAACCCTTGCCTAAGGATTATGGAAGGATTTGTCCAGAGCTCTGAAGACCATGGTGTGGT 702

Qy     924 ggtgttttctctggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgc 983
           ||| |||||||| |||||| || | |||||||||||| ||||  |||| | ||||||
Db     701 GGTCTTTTCTCTTGGGTCAATGGTTAGTAACATGACAGAAGAAAAAGCCAACGCAATTGC 642

Qy     984 aacagcccttgccaagatcccacaaaaggttctgtggagatttgacgggaataaaccaga 1043
           |   |||||||| |||| |||||||||||| |||| |||||| || || |  ||||
Db     641 ATGGGCCCTTGCCCAGATTCCACAAAAGGTTCTTTGGAAATTTGATGGCAAAACCCCAGC 582

Qy    1044 tgccttaggtctcaatactcggctgtacaagtggataccccagaatgaccttctaggtca 1103
           | ||||| | |||||| | |||||||||||| | |||||||||||||| || |||||
Db     581 AACTTTAGGACACAATACCAGAGTGTACAAGTGGCTCCCCCAGAATGACCTCCTTGGTCA 522

Qy    1104 tccaaaaaccagagcttttataactcatggtggagccaatggcatctatgaggcaatcta 1163
           |||||||||||| ||| ||| |||||||||||||| |||||||  |||| ||||| |||||
Db     521 TCCAAAAACCAAAGCCTTTGTAACTCATGGTGGGGCCAATGGTGTCTACGAGGCGATCTA 462

Qy    1164 ccatgggatccctatggtgggcattccattgtttttgatcaacctgataacattgctca 1223
           |||||  |||||||||  | ||||||||  ||||||    || || | ||||||||||| ||
Db     461 TCATGGAATCCCTATGATTGGCATTCCTTTGTTTGGAGACCAGCATGATAACATTGCCCA 402

Qy    1224 catgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagtacagacct 1283
           |||  |||||| |||||||||||  |||| |  |||  ||||||||  |   ||||  |
Db     401 TATGGTGGCCAAAGGAGCAGCTGTTGCATTGAATATCAGAACAATGTCAAAGTCAGATGT 342

Qy    1284 gctgaatgcactgaagacagtaattaatgatcctttatataaagagaatattatgaaatt 1343
           ||| |||||||| ||   || ||  |  ||||||| |||||| |||||  ||||   ||
Db     341 GCTCAATGCACTGGAGGAGGTCATAGAAAATCCTTTCTATAAAAAGAATGCTATGTGGTT 282

Qy    1344 atcaagaattcaacatgatcaaccagtaaagcccctggatcgagcagtcttctggattga 1403
           ||||  |||||  ||||| || ||  | || ||||||||  | || ||||||||||||||
Db     281 GTCAACCATTCACCATGACCAGCCTATGAAACCCCTGGACAGGGCTGTCTTCTGGATTGA 222

Qy    1404 atttgtcatgccccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctg 1463
           ||||| |||| |||||||  |||||| |||||  ||       |  ||| |||| |||||
Db     221 GTTTGTTATGCGCCACAAAAGAGCCAAGCACCTGAGACCACTTGGACATAACCTTACCTG 162

Qy    1464 gttccagtaccactctttggatgtgattgggtttctgctggcctgtgtggcaactgtgat 1523
           || |||||||||||| |||||||||||| || || ||  | ||||||||||    ||
Db     161 GTACCAGTACCACTCTCTGGATGTGATTGGATTCCTACTCTCTTGTGTGGCAACCACTAT 102

Qy    1524 atttatcatcacaaagttttgtctgttttgtttctggaagtttgctagaaaagggaagaa 1583
           | |  | |    |||||      ||||   || | |   ||||  |  ||||  || ||
Db     101 AGTCCTTAGTGTAAAGTGCCTCTTGTTCATTTACCGATTCTTTGTAAAGAAAGAAAATAA 42

Qy    1584 gggaaaaagagattagttatgtctgacatttgaa 1617
              || |  || |||   |  |||||| |  ||
Db      41 AATGAAGAATGAGTAGAGCTCACTGACAATGAAA 8
```

```
RESULT   6
AA255327
LOCUS       AA255327      651 bp    mRNA            EST       14-MAR-1997
DEFINITION  mz84f10.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720139 5'
            similar to gb:U06641 UDP-GLUCURONOSYLTRANSFERASE 2B8 PRECURSOR,
            MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AA255327
VERSION     AA255327.1  GI:1889924
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:445635
```

```
           Seq primer: -28ml3 rev2 ET from Amersham
           High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:720139"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      192 a    144 c    147 g    168 t
ORIGIN


  Query Match             24.8%;  Score 427.4;  DB 4;  Length 651;
  Best Local Similarity   79.7%;  Pred. No. 2.7e-98;
  Matches  517;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy     734 gacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaactcct 793
           | || || |||||| |||||||||||||| ||||| || || ||||| ||  ||  ||  |
Db       1 GGCCAACCACCTTAATTGAGACAATGGGCAAAGCAGAAATGTGGCTCATTAGATCCTATT 60

Qy     794 ggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggattccact 853
           ||  |||  | ||||||||| |||  ||||||||| ||| | | |||||||||| |||| |
Db      61 GGGATTTGGAGTTTCCTCACCCAACCTTACCAAATGTTTACTATGTTGGAGGAGTCCATT 120

Qy     854 gcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctctggagaaa 913
           |||||||||| |||||  | |||||||||||||| || ||||| ||||||||||||||
Db     121 GCAAACCTGCTAAACCATTGCCTAAGGAAATGGAAGACTTTGTTCAGAGCTCTGGAGAGC 180

Qy     914 atggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagggcca 973
           ||||||| |||||||||||||||||||| || | ||||| ||||||| ||||| |||||
Db     181 ATGGTGTGGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAATATGACAGAAGAAAAGGCCA 240

Qy     974 atgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgacggga 1033
           ||| |||||||   || |||||||  || |||||||||||||| |||||||||||||| |
Db     241 ATGCAATTGCATGGGCTCTTGCCAGCATTCCACAAAAGGTTCTTTGGAGATTTGACGGCA 300

Qy    1034 ataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccagaatgacc 1093
           | |  ||||   ||||||| | ||||||  |  | ||||||||| | ||||| ||||| |
Db     301 AAACTCCAGCCACCTTAGGACCCAATACCAGAATCTACAAGTGGCTTCCCCAAAATGATC 360

Qy    1094 ttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatg 1153
           |||| |||||||||||||||| ||| |||||||||||||||| || |||||   |||| |
Db     361 TTCTTGGTCATCCAAAAACCAAAGCCTTTATAACTCATGGTGGCGCGAATGGACTCTACG 420

Qy    1154 aggcaatctaccatgggatccctatggtgggcattccattgttttttgatcaacctgata 1213
           |||| ||| | ||||| ||||||||| | |||||||| ||||||   || |||| |||||
Db     421 AGGCGATCCATCATGGAATCCCTATGATTGGCATTCCTTTGTTTGGAGAACAACATGATA 480

Qy    1214 acattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcga 1273
           |||||||||||||||  |||||| |||||||||||||| |||| |  |||  || ||||| |
Db     481 ACATTGCTCACATGGTGGCCAAAGGAGCAGCTGTTACATTGAATATCAGAACGATGTCAA 540

Qy    1274 gtacagacctgctgaatgcactgaagacagtaattaatgatcctttatataaagagaata 1333
           |  ||||  |||| |||||||| |   ||| ||  |  ||||||| ||||||||||||
Db     541 GGTCAGATTTGCTCAATGCACTGGACGAAGTCATAGACAATCCTTTCTATAAAGAGAATG 600

Qy    1334 ttatgaaattatcaagaattcaacatgatcaaccagtaaagcccctgga 1382
            ||||   || || |  |||||||||| || ||  | ||||||||||
Db     601 CTATGTGGTTGTC-ACCATTCAACATGACCAGCCTATGAAGCCCCTGGA 648


RESULT   7
```

```
AI821970/c
LOCUS       AI821970      669 bp    mRNA            EST       13-DEC-1999
DEFINITION  ob22c12.x5 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1324438 3'
            similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION   AI821970
VERSION     AI821970.1  GI:5441049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ob22c12.y5
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: National Cancer Institute, Cancer Genome
            Anatomy Project (CGAP), Tumor Gene Index
            This read has been verified (found to hit its original self in the
            correct orientation)
            Insert Length: 1566   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1324438"
                     /clone_lib="NCI_CGAP_Kid5"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      189 a    137 c    108 g    235 t
ORIGIN


  Query Match             24.7%;  Score 424.6;  DB 102;  Length 669;
  Best Local Similarity   90.5%;  Pred. No. 1.4e-97;
  Matches  488;  Conservative    0;  Mismatches   44;  Indels    7;  Gaps    3;

Qy    1184 gcattccattgtttttgatcaacctgataacattgctcacatgaaggccaagggagcag 1243
           | ||| |||||||    ||||||| |||||||| |||||||||| ||| || ||||||||
Db     669 GGATTACATTGTTGACCGATCAACTTGATAACA-TGCTCACATG-AGGGCAGGGGAGCAG 612

Qy    1244 ctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacag 1303
           |||||||| |||||||||||| |||||||||||||| |||||||||| |||||| ||
Db     611 CTGTTAGAGTGGACTTCAACACGATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAG 552

Qy    1304 taattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgatc 1363
           ||| |||||||||| |||||||||||||| | ||||||||||||||||||||||||||||
```

```
Db     551 TAAATAATGATCCTTCATATAAAGAGAATGTAATGAAATTATCAAGAATTCAACATGATC 492

Qy    1364 aaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaag 1423
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     491 AACCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAG 432

Qy    1424 gagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttgg 1483
           |||| |||||| |||| |||||||||||| |||||||||||||||||||||||||||||||
Db     431 GAGCTAAACACATTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGG 372

Qy    1484 atgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagtttt 1543
           |||||||||||||  |||||| |||||||||||||||||||||||| ||||||| | ||
Db     371 ATGTGATTGGGTTCATGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTT 312

Qy    1544 gtctgttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttat 1603
           |||||||||||||||||||||||||||||||||  ||||||||||||  ||||||||||
Db     311 GTCTGTTTTGTTTCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTAT 252

Qy    1604 gtctgacatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaag 1663
            ||||| ||||||||||||||||||| ||||| |  ||| ||||||||||||||||||||
Db     251 ATCTGAAATTTGAAGCTGGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAAG 192

Qy    1664 aaagaaaagattgttatgcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaaa 1722
                ||||||||| |||||||||||||||||||||| |||||||||||||||| |||
Db     191 -----AAAGATTGTGATGCAAGATTTCTTTCTTCCTGAGACAAAAAAAAAAAAAAGAAA 138


RESULT   8
AI529976
LOCUS       AI529976      923 bp    mRNA            EST       18-MAR-1999
DEFINITION  ui87g04.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1889430 5' similar to gb:Y00317 UDP-GLUCURONOSYLTRANSFERASE
            2B4 PRECURSOR, MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AI529976
VERSION     AI529976.1  GI:4444111
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:973754
            Seq primer: custom primer used
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1889430"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      273 a    181 c    189 g    277 t      3 others
ORIGIN


  Query Match             24.6%;  Score 423.6;  DB 21;  Length 923;
  Best Local Similarity   69.6%;  Pred. No. 2.8e-97;
  Matches  616;  Conservative    0;  Mismatches  262;  Indels    7;  Gaps    3;

Qy      1 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttac 60
          | || || |||| || ||||| | |||||||| ||||| ||||||| ||| || ||
Db     22 TTCAGCAAGATGTCTATGAAACAGGCTTCAGTTTTTCTGTTGATACAGTTCATATGCTAT 81

Qy     61 tttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatg 120
            ||||  ||||    ||||| |||||||||||||| |  ||||||||||||||||||
Db     82 ATTAGACCTGGAGCCTGTGGGAAAGTGCTGGTGTGGCCTACAGAATACAGCCATTGGATA 141

Qy    121 aatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggca 180
          |||||||| | |||||| | || ||||| |||||||||||||| || || || ||   |
Db    142 AATATGAAAATAATCCTGGATGAACTTGTCCAGAGAGGTCATGACGTCACCGTTCTCATA 201

Qy    181 tcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttat 240
          ||||| |||||||| ||  |||  || |   ||| |||  || |||| |||||| |||||
Db    202 TCTTCTGCTTCCATCCTCATTGGGCCGAGCAATGAATCTTCTATTAATTTTGAAATTTAT 261

Qy    241 cctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg 300
           || || |||| | ||||  |||  ||||  ||    |     ||  |||    | |||
Db    262 TCTGCACCTTTGAGTAAAGATGATCTTGAATATGCTTTTGAAAAATGGGTAGGAAACTGG 321

Qy    301 tca---gacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctg 357
           ||   ||  |  ||||  |  |||||| || || |||| ||   ||| |||||
Db    322 ACATACGAATTAAAAAAACTTCCATTTTGGACATCTTATTCAAAACTGCAAAAAATCTCC 381

Qy    358 tgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagtt 417
          | |||| | |||| || || |||| |||| |||||||  || ||||   |
Db    382 AGTGAATATTCAGACATGATTGAAAGTTTCTGCAAAGCAGTAGTTTGGAACAAGAGCCTC 441

Qy    418 atgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgt 477
          |||||||||| |||  || |  |||||  ||||| | ||||||||  |  ||||||||
Db    442 ATGAAAAAACTCCAAGGATCTAAGTTTGATGTCGTTCTAGCAGATGCCTTGGTTCCCTGT 501

Qy    478 ggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcct 537
          ||||||||||  | |  || ||||| | ||  || || ||||||||||||||   ||||
Db    502 GGTGAGCTGCTATCAGAACTGCTTAAGACACCTTTAGTATACAGTCTCCGCTTCTGTCCT 561

Qy    538 ggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctatt 597
          |  |||| |  ||||| | ||||||| || ||     |||| |||||||  | ||| |
Db    562 GCATACAAATGTGAAAAGTACAGTGGGGGCCTTTCACTCCCCCCTTCCTATGTGCCTGTG 621

Qy    598 gttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctat 657
          ||| ||||| || |||||||| || ||||| ||    ||   ||| || |||||| |  |
Db    622 GTTCTGTCAGAACTAAGTGACCACATGACATTTGCAGAANAGGTGAAGAATATGTTGCAN 681

Qy    658 gtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttac 717
          |||  |   | |||||||||||  | || |  | | |  |||   ||| ||||||||||
Db    682 GTGTTGCTTCTTGACTTTTGGCTTCANACATTTAACGAGAAATCCTGGAATCAGTTTTAC 741

Qy    718 agtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatgg 777
          ||||| ||| |||| ||||| || || |||  ||||  ||||| |||| ||||||||||
Db    742 AGTGATGTTCTAGGGAGACCTACAACATTAACTGAGATGATGGGGAAAGCAGACATATGG 801

Qy    778 cttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgatttt 837
          ||    |||| || |||||    | || || |||| || ||||| || ||  |||| |||
Db    802 CTCGGTCGAACCTTCTGGA--CCTGAAATTTCTCACCCTTTCTTGCCTAATTTTGACTTT 859
```

```
Qy    838 gttggaggattccactgcaaacctgccaaaccctacctaaggaa 882
          | ||||| ||| |  ||| | | |||||| || |  ||| ||
Db    860 GGTGGAGACTTC--ATTGAAAACAGGCAAACCACTGCTAAAGAAA 902


RESULT  9
BF789664
LOCUS       BF789664     1192 bp    mRNA           EST       12-JAN-2001
DEFINITION  602105279F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223443
            5', mRNA sequence.
ACCESSION   BF789664
VERSION     BF789664.1  GI:12094700
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1192)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9811  row: n  column: 20
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1. .1192
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4223443"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      277 a    304 c    276 g    335 t
ORIGIN


  Query Match             24.5%;  Score 421.4;  DB 169;  Length 1192;
  Best Local Similarity   77.3%;  Pred. No. 1.1e-96;
  Matches  537;  Conservative    0;  Mismatches  156;  Indels    2;  Gaps    2;

Qy    826 aacgttgattttgttggaggattccactgcaaacctgccaaaccctacctaaggaaatg 885
          || ||||| | |||||||||| ||||||||||||||||| |||||  | ||||||||| |||
Db      1 AATGTTGACTATGTTGGAGGACTCCACTGCAAACCTGCTAAACCTTTGCCTAAGGATATG 60

Qy    886 gaggagtttgtacagagctctggagaaaatggtgttgtggtgtttttctctggggtcagtg 945
          || || ||||| |||||||||||||  ||||||| |||||||||| || |||||| ||
Db     61 GAAGAATTTGTCCAGAGCTCTGGAGACCATGGTGTGGTGGTGTTTTCACTTGGGTCAATG 120

Qy    946 ataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatccca 1005
           | |||||||||||| |||||  |||||   |||||||   |||||||| |||| |||
Db    121 GTTAGTAACATGACAGAAGAAAAAGCCAACACAATTGCATGGGCCCTTGCCCAGATTCCA 180

Qy   1006 caaaaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaatactcgg 1065
          |||||||||| |||| |||||| || || |  ||||   | ||||| | ||||||  |
Db    181 CAAAAGGTTCTTTGGAAATTTGATGGCAAAACCCCAGCAACTTTAGGACACAATACGAGA 240

Qy   1066 ctgtacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttttata 1125
           |||||||||| | |||||||||||||| || |||||||||||||| | ||| ||| ||
Db    241 GTGTACAAGTGGCTCCCCCAGAATGACCTCCTTGGTCATCCAAAAACAAAAGCCTTTGTA 300

Qy   1126 actcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggc 1185
          |||||||||||| |||||||| |||||||| ||||| ||||| |||||||| | |||
Db     301 ACTCATGGTGGGGCCAATGGTGTCTATGAGGTGATCTATCATGGAATCCCTATGATTGGC 360

Qy   1186 attccattgttttttgatcaacctgataacattgctcacatgaaggccaagggagcagct 1245
          ||||| ||||||   || |||| |||||| ||||| || |||  |||||| |||||||||
Db     361 ATTCCTTTGTTTGGAGAACAACATGATAATATTGCCCATATGGTGGCCAAAGGAGCAGCT 420

Qy   1246 gttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagta 1305
          |||| |||| |  ||| |||||||| ||  |||| | || |||||||||| ||  |||
Db     421 GTTACATTGAATATCAGAACAATGTCAAGATCAGATGTACTCAATGCACTGGAGGAAGTC 480

Qy   1306 attaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaa 1365
          ||  | ||||||| |||||| |||||  |||    || |||| |||| || || ||
Db     481 ATAGACAATCCTTTCTATAAAAAGAATGCTATATGGTTGTCAACCATTCACCACGACCAG 540

Qy   1366 ccagtaaagcccctggatcgagcagtctttctggattgaatttgtca-tgccccacaaagg 1424
          ||    |||||||||||  | || ||||||||| |||| ||||||| ||| ||||||| |
Db     541 CCTACGAAGCCCCTGGACAGGGCTGTCTTCTGGGTTGAGTTTGTCATTGCGCCACAAAAG 600

Qy   1425 agccaaacaccttcgagttgcagccca-tgacctcacctggttccagtaccactctttgg 1483
           || || ||||  ||       |  || | |||| |||||  |||||||| | | | |
Db     601 GGCTAAGCACCTGAGATCACTTGGACATTAACCTTACCTGGCACCAGTACCAATTTCTAG 660

Qy   1484 atgtgattgggtttctgctggcctgtgtggcaact 1518
          ||||||| || || || ||  | ||||| || | |
Db     661 ATGTGATCGGCTTCCTACTCTCTTGTGTTGCCAGT 695


RESULT  10
AA987868/c
LOCUS       AA987868     482 bp    mRNA           EST       02-JUN-1998
DEFINITION  oq81f04.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1592767 3'
            similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION   AA987868
VERSION     AA987868.1  GI:3173232
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1592767"
                     /clone_lib="NCI_CGAP_Kid6"
                     /sex="mixed"
                     /tissue_type="kidney tumor"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
```

GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."
BASE COUNT 136 a 104 c 82 g 160 t
ORIGIN

Query Match 23.8%; Score 410.4; DB 14; Length 482;
Best Local Similarity 92.5%; Pred. No. 5.5e-94;
Matches 444; Conservative 0; Mismatches 31; Indels 5; Gaps 1;

```
Qy   1243 gctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagaca 1302
          ||||||||| | ||||||||||||||||||||||||||| |||||||||| |||||| |
Db    482 GCTGTTAGAGGGTACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGA 423

Qy   1303 gtaattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgat 1362
          |||||||||||||||| |||||||||||| |||||||||||||||||||||||||||||
Db    422 GTAATTAATGATCCTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGAT 363

Qy   1363 caaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaa 1422
          |||||||| ||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    362 CAACCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAA 303

Qy   1423 ggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttg 1482
          ||||| |||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db    302 GGAGCTAAACACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTG 243

Qy   1483 gatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttt 1542
          |||||||||||||| ||||||| |||||||||||||||||||||||| ||||||| | |
Db    242 GATGTGATTGGGTTCCTGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGT 183

Qy   1543 tgtctgttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagtta 1602
          |||||||||||||||||||||||||||||||||  ||||||||||||  |||||||||
Db    182 TGTCTGTTTTGTTTCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTA 123

Qy   1603 tgtctgacatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaa 1662
          | ||||| |||||||||||||||||| ||||| |  ||| ||||||||||||||||||||
Db    122 TATCTGAGATTTGAAGCTGGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAA 63

Qy   1663 gaaagaaaagattgttatgcaagatttctttcttcctgtgacaaaaaaaaaaaaaaaaaa 1722
          |     ||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db     62 G-----AAAGATTGTGATGCAAGATTTCTTTCTTCCTGAGACAAAAAAAAAAAAAAAAA 8
```

RESULT 11
AI530408
LOCUS AI530408 996 bp mRNA EST 18-MAR-1999
DEFINITION ui92f08.y1 Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1889895 5' similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR, MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION AI530408
VERSION AI530408.1 GI:4444543
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 996)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:974219
Seq primer: custom primer used
High quality sequence stop: 527.
FEATURES Location/Qualifiers
source 1. .996
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1889895"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."
BASE COUNT 273 a 180 c 223 g 316 t 4 others
ORIGIN

Query Match 23.8%; Score 409.8; DB 21; Length 996;
Best Local Similarity 66.5%; Pred. No. 9.4e-94;
Matches 630; Conservative 0; Mismatches 311; Indels 6; Gaps 3;

```
Qy      6 caggatgactctgaaatggacttcagttcttctgctgatacatctcagttgttactttag 65
          || |||| ||   || |||| ||| | ||| |||||| | ||  | ||||| | ||| ||
Db     39 CAAGATGCCTGGAAAGTGGATTTCTGCTCTGCTGCTGCTGCAGATAAGTTGCTGCTTCAG 98

Qy     66 ctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatat 125
          |||| ||  ||||| |||||| |||||||| |    |||| ||| ||||||||||||||
Db     99 ATCTGTGAAATGTGGGAAAGTGTTGGTGTGGCCGATGGAATTCAGTCATTGGATGAATAT 158

Qy    126 gaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttc 185
          || | ||| ||| | || || || |||||||| ||||| || ||||| |||  | ||||
Db    159 AAAAATAATACTGGATGAACTGGTACAGAGAGGCCATGAAGTCACTGTTCTGAGACCTTC 218

Qy    186 agcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcctac 245
          ||||| |  | || ||||||||||  |  || |    ||||| ||||||  || ||||||
Db    219 AGCTTACTATGTTCTTGATCCCAAAAAATCACCTGGCCTTAAGTTTGAAACTTTTCCTAC 278

Qy    246 atctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg---tc 302
          |||| | | ||||    |. | || ||| |  ||||  |    ||  |   |||   |
Db    279 ATCTGTCAGTAAAGACAATCTGGAAAATTTTTTCATTAAGTTTGTGGATGTGTGGACTTA 338

Qy    303 agacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtggga 362
          || || | ||  |||| | || || ||||| | |   |   ||| | ||  |    ||
Db    339 TGAAATGCCAAGGGATACATGTTTGTCATATTCTCCTTTACTGCAAAATATGATTGATGA 398

Qy    363 attatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaa 422
          ||| | |||   |||  ||   | |||||||||| |||||||| ||  |  | ||||
Db    399 ATTTTCTGATTATTTTCTAAGTCTTTGTAAAGATGTTGTTTCAAACAAAGAGCTCATGAC 458

Qy    423 aaaactacaagagtcaagatttgacatcgtttttgcagatgctgttttttccctgtggtga 482
          ||| ||||| || || |  |||||  || ||||  |||||  |||||  |  |||||| ||
Db    459 AAAGCTACAGGAATCCAAGTTTGATGTCCTTTTGTCAGATCCTGTTGCTTTCTGTGGNGA 518

Qy    483 gctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggcta 542
          |||  | ||||  || ||  | || |  ||| |||||||| |||||||  |||| |||||
Db    519 GCTCATAGCTGAACTGCTCCAGATTCCTTTTCTGTACAGTATCCGCTTCTCTCCAGGCTA 578

Qy    543 cacaattgaaaggcacagtggaggactgattttcctccttcctacatacctattgttat 602
          |  ||||||||    ||| ||  || |  |  ||| ||| || ||  ||||| |  || |
Db    579 CCAAATTGAAAAATCCAGCGGGAGATTTCTACTCCGTCCCTCTTATGTACCTGTNATTCT 638
```

```
Qy    603 gtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtgct 662
          ||||  |||| |||  |||||||| ||||| |||||| |||||||||||||| |   ||||
Db    639 GTCAGGATTAGGTGGACAAATGACATTCATAGAGAGGATAAAAAATATGATATGCATGCT 698

Qy    663 ttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagtga 722
          || |||||||| | ||||| |||| | |     |||| |||||    ||||||||||
Db    699 NTACTTTGACTTTGGCTTCCAGATGTTTAAATGATAGAAATGGGACTCATTTTACAGTGA 758

Qy    723 agttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttat 782
          |||  ||||| || || |||||| ||||||| |||||  ||||    || ||||||
Db    759 GTATTTGGGAAGGCCAACCACCTTAATTGAGACCATGGGCCAAGCNAGAATGTGGCTTCA 818

Qy    783 gcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttgg 842
          ||| |     | | |||  | || |||| |||  |||  | || || || | || |||
Db    819 TAGAATCAAACTGGGATTTGGAGTTTCTCACCCAACCTTTACCAATGTGGACTATGGTGG 878

Qy    843 aggattccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagag 902
          |||||  |   |  |||||||  |||||  |||||||  ||  ||  |||| ||||
Db    879 ACGATGC--ATTGCAACCTGCTTAACCCTGGCCTAAGGGATAGGGAGA-ATTGTCCAGAC 935

Qy    903 ctctggagaaaatggtgttgtggtgttttctctggggtcagtgataa 949
          ||||  | |   ||||       |  |||    |||     ||  |||
Db    936 TTCTGAAAACCCTGGTTGGTGTGCTTTTCTCTTGGTCAATGTTGTAA 982


RESULT 12
AA255225
LOCUS       AA255225      616 bp    mRNA            EST       14-MAR-1997
DEFINITION  mz85d04.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720199 5'
            similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR,
            MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AA255225
VERSION     AA255225.1  GI:1889794
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:445695
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:720199"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      179 a    124 c    140 g    173 t
ORIGIN


  Query Match           23.7%;  Score 408.2;  DB 4;  Length 616;
  Best Local Similarity   79.9%;  Pred. No. 2.1e-93;
  Matches  493;  Conservative    0;  Mismatches  123;  Indels    1;  Gaps    1;

Qy    529 tttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctac 588
          ||  ||||||||||    ||||||   |||||||||||||    | ||||||| || ||
Db      1 TTCTCTCCTGGCTACTATCTTGAAAAATACAGTGGAGGACTCCCTCTCCCTCCCTCTTAT 60

Qy    589 atacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaat 648
          |||||  |  || |||||  |||||||| |||||||| ||||  |||||||||||  |||
Db     61 GTACCTGTAATTTTGTCAGGATTAAGTGGTCAAATGACATTCAAAGAGAGGGTAAAGAAT 120

Qy    649 atgatctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggat 708
          ||||| | | |||||||||||||||||||||||| |  | |     |||||| ||||||
Db    121 ATGATATGTATGCTTTATTTTGACTTTTGGTTCCAGACATTTAGGGAGAAGAAATGGGAT 180

Qy    709 cagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagct 768
          ||||||||||| |||  ||| ||||| || || |||||| ||||||||||||| |||||
Db    181 CAGTTTTACAGCGAAACTTTGGGAAGGCCAACCACCTTAATTGAGACAATGGGCAAAGCA 240

Qy    769 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 828
          || || ||||| ||  ||  ||  ||| |||  | ||||||| |||  |||||||||
Db    241 GAAATGTGGCTCATTAGATCCTATTGGGATTTGGAGTTTCCTCACCCAACCTTACCAAAT 300

Qy    829 gttgattttgttggaggattccactgcaaacctgccaaacccctacctaaggaaatggag 888
          ||| | | ||||||||||| |||| |||||||||| |||||  | ||||||||||||||
Db    301 GTTTACTATGTTGGAGGAGTCCATTGCAAACCTGCTAAACCATTGCCTAAGGAAATGGAA 360

Qy    889 gagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgata 948
          || ||||| |||||||||||||  ||||||| |||||||||||||||||| | || |
Db    361 GACTTTGTTCAGAGCTCTGGAGAGCATGGTGTGGTGGTGTTTTCTCTGGGGTAAATGGTC 420

Qy    949 agtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaa 1008
          ||||| ||||||| ||||| |||||||| |||||||   || |||||| |||| ||||||
Db    421 AGTAATATGACAGAAGAAAAGGCCAATGCAATTGCATGGGCTCTTGCCCAGATTCCACAA 480

Qy   1009 aaggttctgtggagatttgacgggaataaaccagatgccttaggtctcaatactcggctg 1068
          |||||||| |||||||||||||| || |  ||||   |||||||  | ||||||  |  |
Db    481 AAGGTTCTTTGGAGATTTGACGGCAAAACTCCAGCCACCTTAGGACCCAATACCAGAATC 540

Qy   1069 tacaagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataact 1128
          |||||||| | ||||| ||||| ||||||| |||||||  ||  |||||||||
Db    541 TACAAGTGGCTTCCCCAAAATGATCTTCTTGGTCATCC-AAAACCAAAGGCTTTATAACT 599

Qy   1129 catggtggagccaatgg 1145
          |||||||| || |||||
Db    600 CATGGTGGCGCGAATGG 616


RESULT 13
AW604423/c
LOCUS       AW604423      517 bp    mRNA            EST       23-MAR-2000
DEFINITION  RC3-CT0254-200100-024-g07 CT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW604423
VERSION     AW604423.1  GI:7309177
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
```

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-CT0254-200100-024-g07&t3=2000-01-20&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 517.

FEATURES Location/Qualifiers
source 1. .517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0254"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 170 a 119 c 91 g 137 t
ORIGIN

Query Match 23.1%; Score 397.6; DB 117; Length 517;
Best Local Similarity 86.4%; Pred. No. 1e-90;
Matches 439; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy    466 gtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctc 525
          |||  |||||||||||||||||||||||| |||||||||||||||  ||| ||||||||||
Db    516 GTTAATCCCTGTGGTGAGCTGCTGGCTGAGCTACTTAACATACCCTTTCTGTACAGTCTC 457

Qy    526 cgctttactcctggctacacaattgaaaggcacagtggaggactgatttttccctccttcc 585
          |||||  ||  |||||||||| |||| | | |  |||||||| |  | ||||||||||||
Db    456 CGCTTCTCTGTTGGCTACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCC 397

Qy    586 tacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaa 645
          ||  ||||| |||||||||| |||||||||||||||||| |||||||||||| |||||
Db    396 TATGTACCTGTTGTTATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAA 337

Qy    646 aatatgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgg 705
          |||||||| ||| |||||||||||||||||||||| |||   | |||| ||||||||||
Db    336 AATATGATATATATGCTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGG 277

Qy    706 gatcagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaa 765
          || |||||||| ||||||||| |||||||||||||||| |||||||||||||||||| |||
Db    276 GACCAGTTTTATAGTGAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAA 217

Qy    766 gctgacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttacca 825
          ||||| || ||||| || |||| ||  ||   |||| |||||||||  ||||||||||||
Db    216 GCTGAAATGTGGCTCATTCGAACCTATTGTGATTTTGAATTTCCTCGCCCATTCTTACCA 157

Qy    826 aacgttgattttgttggaggattccactgcaaacctgccaaaccccctacctaaggaaatg 885
          || |||||||||||||| | ||||| ||||| |||||||| | ||| ||||||||
Db    156 AATGTTGATTTTGTTGGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTTAGGAAATG 97

Qy    886 gaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtg 945
          || |||||||| ||||||||||||||||||||| ||||||||||||||||||||||  ||
Db     96 GAAGAGTTTGTGCAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATG 37

Qy    946 ataagtaacatgacagcagaaagggcca 973
          || ||||||||| ||| |||||| ||||
Db     36 ATCAGTAACATGTCAGAAGAAAGTGCCA 9
```

RESULT 14

BF786091
LOCUS BF786091 792 bp mRNA EST 12-JAN-2001
DEFINITION 602110662F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4239110 5', mRNA sequence.
ACCESSION BF786091
VERSION BF786091.1 GI:12091127
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 792)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9851 row: k column: 15
High quality sequence start: 3
High quality sequence stop: 705.

FEATURES Location/Qualifiers
source 1. .792
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4239110"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT 216 a 165 c 186 g 225 t
ORIGIN

Query Match 23.0%; Score 396; DB 169; Length 792;
Best Local Similarity 74.9%; Pred. No. 2.9e-90;
Matches 535; Conservative 0; Mismatches 175; Indels 4; Gaps 3;

```
Qy    471 tccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctt 530
          | ||||||| |||||  | ||||  || ||  | || |  ||| |||||||| |||||||
Db     41 TTCCTGTGGGGAGCTCATAGCTGAACTGCTCCAGATTCCTTTTCTGTACAGTATCCGCTT 100

Qy    531 tactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacat 590
            |||| ||||||  |||||||| |  ||| ||  ||  |  |  ||||||| || ||  |
Db    101 CTCTCCAGGCTACCAAATTGAAAAGTCCAGCGGGAGATTTCTACTCCCTCCCTCTTATGT 160

Qy    591 acctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaaatat 650
          |||| |  || |||||  |||| |||  |||||||| ||||| ||||| ||||||||||
Db    161 ACCTGTAATTCTGTCAGGATTAGGTGGACAAATGACATTCATAGAGAGAGTAAAAAATAT 220

Qy    651 gatctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatca 710
          ||| |    |||||||||     |||||||||| |||| | ||   ||||| |||||
Db    221 GATATGCAGGCTTTATTT--GACTTTGGTTCCAGATGTTTAATGATAAGAAATGGGACTC 278

Qy    711 gttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctga 770
          ||||||||||||  ||| ||||| ||||| ||||||  |||||||||||| || || ||
Db    279 ATTTTACAGTGAATATTTGGGAAGGCCCACCACCTTAGCTGAGACAATGGGCAAGGCAGA 338

Qy    771 catatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgt 830
          || ||||| ||  ||  |  |||| |||  | ||||||| |||  |||||||| ||
Db    339 AATGTGGCTCATTAGATCCAACTGGGATTTGGAGTTTCCTCACCCAACCTTACCAAATGT 398

Qy    831 tgattttgttggaggattccactgcaaacctgccaaacccctacctaaggaaatggagga 890
```

```
          ||| | |||||||||| |||||||||||||||| |||||  | |||||||| ||||| ||
Db    399 TGACTATGTTGGAGGACTCCACTGCAAACCTGCTAAACCTTTGCCTAAGGATATGGAAGA 458

Qy    891 gtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataag 950
          |||||| |||||||||||||  ||||||| |||||||||| || |||||| || | ||
Db    459 ATTTGTCCAGAGCTCTGGAGACCATGGTGTGGTGGTGTTTTCACTTGGGTCAATGGTTAG 518

Qy    951 taacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaa 1010
          |||||||||| |||||  |||||   |||||||   |||||||| |||| ||||||||
Db    519 TAACATGACAGAAGAAAAAGCCAACACAATTGCATGGGCCCTTGCCCAGATTCCACAAAA 578

Qy   1011 ggttctgtggagatttgacgggaataaaccagatgccttaggtctcaatactcggctgta 1070
          |||||| |||| |||||| || || |  ||||   | ||||| | ||||||  |  ||||
Db    579 GGTTCT-TGGAAATTTGATGGCAAAACCCCAGCAACTTTAGGACACAATACGAGAGTGTA 637

Qy   1071 caagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactca 1130
          ||||||| | |||||||||||||| || |||||||||||| || |    ||   |||||
Db    638 CAAGTGG-TCCCCCAGAATGACCTCCTTGGTCATCCAAAAAACAAAAGCCTTTGTACTCA 696

Qy   1131 tggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtggg 1184
          |||||| ||||  |          ||  ||||| ||||||||||      ||||
Db    697 TGGTGGGGCCATGGTGTCTATGAGGGTGATCTATCATGGGATCCTATGATTGGG 750


RESULT 15
AV657530
LOCUS       AV657530      628 bp    mRNA            EST       07-SEP-2000
DEFINITION  AV657530 GLC Homo sapiens cDNA clone GLCFCH04 3', mRNA sequence.
ACCESSION   AV657530
VERSION     AV657530.1  GI:9878544
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
            Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA clone
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCFCH04"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      189 a    128 c    149 g    162 t
ORIGIN


  Query Match           22.7%;  Score 390.4;  DB 31;  Length 628;
  Best Local Similarity  84.9%;  Pred. No. 7.4e-89;
  Matches  472;  Conservative    0;  Mismatches   31;  Indels   53;  Gaps    1;

Qy    727 ttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcga 786
          | |||||||||||||| |||| |||||||||| | ||||||||||||||||||||||
Db    126 TCAGGAAGACCCACTACATTATCTGAGACAATGAGGAAAGCTGACATATGGCTTATGCGA 185

Qy    787 aactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggagga 846
          |||||||||| |||| ||||||||||||||||||||||||| ||||| ||||||||||
Db    186 AACTCCTGGAATTTTAAATTTCCTCATCCATTCTTACCAAATGTTGAATTTGTTGGAGGA 245

Qy    847 ttccactgcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctct 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTACAGAGCTCT 305

Qy    907 ggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaa 966
          |||||||||||||||||||| |||||||||||| || | |||||||||||| ||||
Db    306 GGAGAAAATGGTGTTGTGGTGTATTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAA 365

Qy    967 agggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagattt 1026
          |||||||| |||||||||||||||||||||||||||||||||| ||| ||||||| |
Db    366 AGGGCCAACGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGGTCTTTGGAGATAT 425

Qy   1027 gacgggaataaaccagatgccttaggtctcaatactcggctgtacaagtggataccccag 1086
          || |||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    426 GATGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGACTGTACAAGTGGATACCCCAG 485

Qy   1087 aatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggc 1146
          |||||||||||||||||| ||
Db    486 AATGACCTTCTAGGTCATTCA---------------------------------------- 506

Qy   1147 atctatgaggcaatctaccatgggatccctatggtgggcattccattgtttttgatcaa 1206
                     |||||||||||||||||||||||||||||||||||||||| || |||||||
Db    507 -------------CTACCATGGGATCCCTATGGTGGGCATTCCATTGTGTTATGATCAA 552

Qy   1207 cctgataacattgctcacatgaaggccaagggagcagctgttagattggacttcaacaca 1266
          |||||||| |||||||||||||||||||||||| |   |||| ||||||| ||||||
Db    553 CCTGATAATATTGCTCACATGAAGGCCAAGGGAGCACCATGTAGAGTGGACTTTAACACA 612

Qy   1267 atgtcgagtacagacc 1282
          ||||||||||||||||
Db    613 ATGTCGAGTACAGACC 628


Search completed: April 30, 2001, 12:18:01
Job time: 12550 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 18:54:38 ; Search time 22667.1 Seconds
(without alignments)
465.367 Million cell updates/sec

Title: US-09-721-183-2
Perfect score: 1713
Sequence: 1 atcgcattgcaccaggatga.....gatttctttcttcctgtgac 1713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1675.6	97.8	1722	88	AF016492	AF016492 Homo sapi
2	1452.4	84.8	2799	93	HSUGT2BIO	X63359 H.sapiens U
3	1450	84.6	1855	93	HUMUDPGTA	J05428 Human 3,4-c
4	1425.8	83.2	1753	88	AF016310	AF016310 Macaca fa
5	1419.4	82.9	2648	93	MFU91582	U91582 Macaca fasc
6	1411.4	82.4	1768	88	AF112113	AF112113 Macaca fa
7	1355.2	79.1	1662	88	AF294901	AF294901 Macaca mu
8	1354.2	79.1	2079	92	HSAJ5162	AJ005162 Homo sapi
9	1353.6	79.0	1662	88	AF294902	AF294902 Macaca mu
10	1351	78.9	2097	88	AF064200	AF064200 Homo sapi
11	1344.6	78.5	2091	88	AF081793	AF081793 Homo sapi

	12	1343	78.4	2093	93	HSUDPGT	Y00317 Human mRNA
	13	1329.8	77.6	2108	88	AF112112	AF112112 Macaca fa
	14	1221.4	71.3	1650	9	AX040087	AX040087 Sequence
	15	1201.6	70.1	2107	93	HSU59209	U59209 Homo sapien
	16	1190.8	69.5	2114	88	AF180322	AF180322 Homo sapi
	17	1188.8	69.4	2090	93	HSU08854	U08854 Human UDP g
	18	1154.6	67.4	2123	93	HSU06641	U06641 Human UDP g
	19	1152.6	67.3	2075	88	AF072223	AF072223 Macaca fa
	20	1004	58.6	1671	7	RABUGT2BB	L01082 Oryctolagus
	21	975	56.9	2818	7	OCU72742	U72742 Oryctolagus
	22	966.2	56.4	1832	7	RABUGT2BA	L01081 Oryctolagus
	23	942.8	55.0	1961	94	RATUDPGTP	M13506 Rat liver U
	24	873.8	51.0	1858	94	MMUDPGT	X06358 Mouse mRNA
	25	870.2	50.8	1844	95	RNU06273	U06273 Rattus norv
	26	869.6	50.8	1716	95	RNUDPGT	Y00156 Rat mRNA fo
	27	864.8	50.5	1819	94	RATGRT	M31109 Rat UDP-glu
	28	863.8	50.4	1846	95	RNU06274	U06274 Rattus norv
	29	845	49.3	1858	94	RATUDPA	M33746 Rat UDP glu
	30	833	48.6	2634	94	RATUDPGTR	J02589 Rat UDP glu
	31	832.6	48.6	1947	95	RNU27518	U27518 Rattus norv
	32	807	47.1	1846	95	RNUDPGTR	X03478 Rat liver m
	33	766	44.7	2547	94	AF175221	AF175221 Cavia por
	34	751.2	43.9	1766	92	HSA6054	AJ006054 Homo sapi
	35	730	42.6	1584	94	AF184901	AF184901 Mus muscu
	36	697.4	40.7	1831	95	RRUGTGENE	X57565 Rattus norv
	37	697	40.7	2897	7	RABUGT2BC	L01083 Oryctolagus
	38	691.8	40.4	2786	89	AK025587	AK025587 Homo sapi
	39	684.8	40.0	897	88	AF179877	AF179877 Homo Sapi
c	40	673	39.3	139015	69	AC025488	AC025488 Homo sapi
	41	671.6	39.2	1021	88	AF179876	AF179876 Homo Sapi
	42	633.4	37.0	138665	72	AC055794	AC055794 Homo sapi
	43	633.4	37.0	212904	66	AC021146	AC021146 Homo sapi
	44	581.6	34.0	1007	88	AF179879	AF179879 Homo sapi
c	45	579.2	33.8	27310	88	AF135416	AF135416 Homo sapi

ALIGNMENTS

```
RESULT  1
AF016492
LOCUS       AF016492     1722 bp    mRNA            PRI       31-JUL-1998
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 2B mRNA, complete cds.
ACCESSION   AF016492
VERSION     AF016492.1  GI:3360272
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1722)
  AUTHORS   Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
  TITLE     Isolation and characterization of a human orphan
            UDP-glucuronosyltransferase, UGT2B11
  JOURNAL   Biochem. Biophys. Res. Commun. 248 (1), 44-50 (1998)
  MEDLINE   98340847
REFERENCE   2  (bases 1 to 1722)
  AUTHORS   Levesque,E., Beaulieu,M., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) Molecular Endocrinology, Laval University,
            2705 Laurier Blvd., Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .1722
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             10. .1599
                     /note="LN117"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B"
                     /protein_id="AAC27891.1"
                     /db_xref="GI:3360273"
                     /translation="MTLKWTSVLLLIHLSCYFSSGSCGKVLVWAAEYSHWMNMKTILK
                     ELVQRGHEVTVLASSASILFDPNDASTLKFEVYPTSLTKTEFENIIMQQVKRWSDIRK
                     DSFWLYFSQEQEILWELYDIFRNFCKDVVSNKKVMKKLQESRFDIVFADAVFPCGELL
                     AALLNIRFVYSLRFTPGYTIERHSGGLIFPPSYIPIVMSKLSDQMTFMERVKNMIYVL
                     YFDFWFQMSDMKKWDQFYSEVLGRPTTLFETMGKADIWLMRNSWSFQFPHPFLPNVDF
                     VGGFHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSVISNMTAERANVIATALAKIPQK
                     VLWRFDGNKPDALGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     IPLFFDQPDNIAHMKAKGAAVRLDFNTMSSTDLLNALKTVINDPLYKENIMKLSRIQH
                     DQPVKPLDRAVFWIEFVMPHKGAKHLRVAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     TKFCLFCFWKFARKGKKGKRD"
BASE COUNT      523 a    319 c    371 g    509 t
ORIGIN
```

```
  Query Match             97.8%;  Score 1675.6;  DB 88;  Length 1722;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1700;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy       8 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccagttgtta 67
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       1 TGCACCAGGATGACTCTGAAATGGACTTCAGTTCTTCTGCTGATACATCT-CAGTTGTTA 59

Qy      68 ctttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggat 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 CTTTAGCTCTGGGAGTTGTGGAAAAGTGCTGGTGTGGGCCGCAGAATACAGCCATTGGAT 119

Qy     128 gaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggc 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 GAATATGAAGACAATCCTGAAAGAGCTTGTTCAGAGAGGTCATGAGGTGACTGTACTGGC 179

Qy     188 atcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagttta 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 ATCTTCAGCTTCCATTCTTTTTGATCCCAATGATGCATCCACTCTTAAATTTGAAGTTTA 239

Qy     248 tcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatg 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 TCCTACATCTTTAACTAAAACTGAATTTGAGAATATCATCATGCAACAGGTTAAGAGATG 299

Qy     308 gtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtg 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GTCAGACATTCGAAAAGATAGCTTTTGGTTATATTTTTCACAAGAACAAGAAATCCTGTG 359

Qy     368 ggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttat 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 GGAATTATATGACATATTTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAAGTTAT 419

Qy     428 gaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtgg 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 GAAAAAACTACAAGAGTCAAGATTTGACATCGTTTTTGCAGATGCTGTTTTTCCCTGTGG 479

Qy     488 tgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctgg 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 TGAGCTGCTGGCTGCGCTACTTAACATACGGTTTGTGTACAGTCTCCGCTTTACTCCTGG 539

Qy     548 ctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgt 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 CTACACAATTGAAAGGCACAGTGGAGGACTGATTTTCCCTCCTTCCTACATACCTATTGT 599

Qy     608 tatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgt 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 TATGTCAAAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGT 659

Qy     668 gctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacag 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 GCTTTATTTTGACTTTTGGTTCCAAATGTCTGATATGAAGAAGTGGGATCAGTTTTACAG 719

Qy     728 tgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggct 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 TGAAGTTTTAGGAAGACCCACTACCTTATTTGAGACAATGGGAAAAGCTGACATATGGCT 779

Qy     788 tatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgt 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 TATGCGAAACTCCTGGAGTTTTCAATTTCCTCATCCATTCTTACCAAACGTTGATTTTGT 839

Qy     848 tggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggagtttgtac 907
```

```
Db    840 TGGAGGATTCCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTAC 898

Qy    908 agagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatga 967
Db    899 AGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAGTGATAAGTAACATGA 958

Qy    968 cagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgt 1027
Db    959 CAGCAGAAAGGGCCAATGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGTTCTGT 1018

Qy   1028 ggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataagtgga 1087
Db   1019 GGAGATTTGACGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGGCTGTACAAGTGGA 1078

Qy   1088 taccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggag 1147
Db   1079 TACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAG 1138

Qy   1148 ccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttt 1207
Db   1139 CCAATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGGTGGGCATTCCATTGTTTT 1198

Qy   1208 gggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggact 1267
Db   1199 TTGATCAACCTGATAACATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGATTGGACT 1258

Qy   1268 tcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctt 1327
Db   1259 TCAACACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTT 1318

Qy   1328 tatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccc 1387
Db   1319 TATATAAAGAGAATATTATGAAATTATCAAGAATTCAACATGATCAACCAGTAAAGCCCC 1378

Qy   1388 tggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttc 1447
Db   1379 TGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCCCCACAAAGGAGCCAAACACCTTC 1438

Qy   1448 gagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttc 1507
Db   1439 GAGTTGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTTC 1498

Qy   1508 tgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttct 1567
Db   1499 TGCTGGCCTGTGTGGCAACTGTGATATTTATCATCACAAAGTTTTGTCTGTTTTGTTTCT 1558

Qy   1568 ggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaag 1627
Db   1559 GGAAGTTTGCTAGAAAAGGGAAGAAGGGAAAAAGAGATTACTTATGTCTGACATTTGAAG 1618

Qy   1628 ctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtt 1687
Db   1619 CTGGAAAACCAGATAGATAGGACAACTTCAGTTTATTCCAGCAAGAAAGAAAAGATTGTT 1678

Qy   1688 atgcaagatttctttcttcctgtgac 1713
Db   1679 ATGCAAGATTTCTTTCTTCCTGTGAC 1704
```

```
RESULT   2
HSUGT2BIO
LOCUS       HSUGT2BIO     2799 bp    mRNA            PRI       19-JUL-1994
DEFINITION  H.sapiens UGT2BIO mRNA for udp glucuronosyltransferase.
ACCESSION   X63359
VERSION     X63359.1  GI:516149
KEYWORDS    UDP-glucuronosyltransferase; ugt2bio gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2799)
  AUTHORS   Mackenzie,P.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1991) P.I. Mackenzie, Flinders University, Dept
            of Clinical Pharmacology, School of Medicine, Flinders Medical
            Centre, Bedford Park SA 5042, AUSTRALIA
REFERENCE   2  (bases 1 to 2799)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
  MEDLINE   93326164
FEATURES             Location/Qualifiers
     source          1. .2799
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11"
                     /clone="UGTh-46"
     gene            11. .1597
                     /gene="UGT2 BIO"
     CDS             11. .1597
                     /gene="UGT2 BIO"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="CAA44961.1"
                     /db_xref="GI:516150"
                     /db_xref="SWISS-PROT:P36537"
                     /translation="MALKWTTVLLIQLSFYFSSGSCGKVLVWAAEYSLWMNMKTILKE
                     LVQRGHEVTVLASSASILFDPNDSSTLKLEVYPTSLTKTEFENIIMQLVKRLSEIQKD
                     TFWLPFSQEQEILWAINDIIRNFCKDVVSNKKLMKKLQESRFDIVFADAYLPCGELLA
                     ELFNIPFVYSHSFSPGYSFERHSGGFIFPPSYVPVVMSKLSDQMTFMERVKNMLYVLY
                     FDFWFQIFNMKKWDQFYSEVLGRPTTLSETMRKADIWLMRNSWNFKFPHPFLPNVDFV
                     GGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNMTEERANVIATALAKIPQKV
                     LWRFDGNKPDALGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGI
                     PLFFDQPDNIAHMKAKGAAVRVDFNTMSSTDLLNALKTVINDPSYKENIMKLSRIQHD
                     QPVKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWFQYHSLDVIGFLLACVATVLFIIT
                     KCCLFCFWKFARKGKKGKRD"
     polyA_signal    2772. .2788
BASE COUNT      792 a    605 c    569 g    833 t
ORIGIN
```

```
  Query Match           84.8%;  Score 1452.4;  DB 93;  Length 2799;
  Best Local Similarity  92.2%;  Pred. No. 0;
  Matches 1553;  Conservative    0;  Mismatches  126;  Indels    5;  Gaps    2;

Qy      8 tgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccagttgtta 67
Db      2 TGCACAAGGATGGCTCTGAAATGGACTACAGTTCTGCTGATACAACT----CAGTTTTTA 57

Qy     68 ctttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggat 127
Db     58 CTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTATGGGCCGCAGAATACAGCCTTTGGAT 117

Qy    128 gaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggc 187
Db    118 GAATATGAAGACAATCCTGAAAGAACTTGTTCAGAGAGGTCATGAGGTGACTGTACTGGC 177

Qy    188 atcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagttta 247
Db    178 ATCTTCAGCTTCCATTCTTTTTGATCCCAACGACTCATCCACTCTTAAACTTGAAGTTTA 237

Qy    248 tcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatg 307
Db    238 TCCTACATCTTTAACTAAAACTGAATTTGAGAATATCATCATGCAATTGGTTAAGAGATT 297

Qy    308 gtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtg 367
Db    298 GTCAGAAATTCAAAAAGATACATTTTGGTTACCTTTTTCACAAGAACAAGAAATCCTGTG 357

Qy    368 ggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttat 427
```

```
Db    358 GGCAATTAATGACATAATTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAACTTAT 417

Qy    428 gaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtgg 487
          ||||||||||||||||||||||||||||||||||||||||||||||||  ||| ||||||||
Db    418 GAAAAAACTACAAGAGTCAAGATTTGACATCGTTTTTGCAGATGCTTATTTACCCTGTGG 477

Qy    488 tgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctgg 547
          |||||||||||||| |||| |||||||||  |||||||||||||| | |||| | ||||||
Db    478 TGAGCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCACAGCTTCAGTCCTGG 537

Qy    548 ctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgt 607
          |||| || |||||||||||||||||||| | |||||||||||||||||| ||||| ||||
Db    538 CTACTCATTTGAAAGGCACAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTGTTGT 597

Qy    608 tatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgt 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    598 TATGTCAAAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGCTCTATGT 657

Qy    668 gctttatttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacag 727
          |||||||||||||||||||||||||| | | ||||||||||||||||||||||||||||
Db    658 GCTTTATTTTGACTTTTGGTTCCAAATATTTAATATGAAGAAGTGGGATCAGTTTTACAG 717

Qy    728 tgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggct 787
          |||||||||||||||||||||||| |||| |||||||||| | ||||||||||||||||||
Db    718 TGAAGTTTTAGGAAGACCCACTACATTATCTGAGACAATGAGGAAAGCTGACATATGGCT 777

Qy    788 tatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgt 847
          |||||||||||||||| |||| ||||||||||||||||||||||||||||||||||||||
Db    778 TATGCGAAACTCCTGGAATTTTAAATTTCCTCATCCATTCTTACCAAATGTTGATTTTGT 837

Qy    848 tggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggagtttgtac 907
          ||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    838 TGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTAC 896

Qy    908 agagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatga 967
          |||||||||||||||||||||||||||||||||||||||||||| || | ||||||||||
Db    897 AGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGA 956

Qy    968 cagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgt 1027
          ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||| |
Db    957 CAGAAGAAAGGGCCAACGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGTTCTTT 1016

Qy   1028 ggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataagtgga 1087
          |||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||
Db   1017 GGAGATTTGATGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGACTGTACAAGTGGA 1076

Qy   1088 taccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggag 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1077 TACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAG 1136

Qy   1148 ccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttt 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1137 CCAATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGGTGGGCATTCCATTGTTTT 1196

Qy   1208 gggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggact 1267
            |||||||||||||| ||||||||||||||||||||||||||||||||| ||||||
Db   1197 TTGATCAACCTGATAATATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGAGTGGACT 1256

Qy   1268 tcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctt 1327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1257 TCAACACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTT 1316

Qy   1328 tatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccc 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   1317 CATATAAAGAGAATATTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCC 1376

Qy   1388 tggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttc 1447
          |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||
Db   1377 TGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCCAAACATCTTC 1436

Qy   1448 gagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttc 1507
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||| |
Db   1437 GAGTTGCAGCCCACAACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCC 1496

Qy   1508 tgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttct 1567
          ||||||| |||||||||| ||| ||||||||||||||||| ||||||||||||||||||
Db   1497 TGCTGGCTTGTGTGGCAACCGTGCTATTTATCATCACAAAGTGTTGTCTGTTTTGTTTCT 1556

Qy   1568 ggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaag 1627
          |||||||||||||||||| |||||||||||||| |||||||||| ||||| ||||||||
Db   1557 GGAAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGGGATTAGTTATATCTGAGATTTGAAG 1616

Qy   1628 ctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtt 1687
          |||| ||     |  || | | |       |     |  |||| |    ||  |
Db   1617 CTGGGGAATTCCGTTTATTGAAGATTCAGGTTAACCTGAATCAAGTTAACCCAGTCTCAA 1676

Qy   1688 atgc 1691
          ||||
Db   1677 ATGC 1680


RESULT   3
HUMUDPGTA
LOCUS       HUMUDPGTA    1855 bp    mRNA            PRI       03-AUG-1993
DEFINITION  Human 3,4-catechol estrogen UDP-glucuronosyltransferase mRNA,
            complete cds.
ACCESSION   J05428
VERSION     J05428.1  GI:340079
KEYWORDS    3,4-catechol estrogen UDP-glucuronosyltransferase.
SOURCE      Human liver, cDNA to mRNA, clone 63-11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1855)
  AUTHORS   Ritter,J.K., Sheen,Y.Y. and Owens,I.S.
  TITLE     Cloning and expression of human liver UDP-glucuronosyltransferase
            in COS-1 cells: 3,4-catechol estrogens and estriol as primary
            substrates
  JOURNAL   J. Biol. Chem. 265, 7900-7906 (1990)
  MEDLINE   90243659
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by I.S.Owens, 22-FEB-1990.
FEATURES             Location/Qualifiers
     source          1. .1855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             15. .1604
                     /note="UDP-glucuronosyltransferase (EC 2.4.1.17)"
                     /codon_start=1
                     /protein_id="AAA36793.1"
                     /db_xref="GI:340080"
                     /translation="MSVKWTSVILLIQLSFCFSSGNCGKVLVWAAEYSHWMNIKTILD
                     ELIQRGHEVTVLASSASILFDPNNSSALKIEIYPTSLTKTELENFIMQQIKRWSDLPK
                     DTFWLYFSQVQEIMSIFGDITRKFCKDVVSNKKFMKKVQESRFDVIFADAIFPCSELL
                     AELFNIPFVYSLSFSPGYTFEKHSGGFIFPPSYVPVVMSELTDQMTFMERVKNMIYVL
                     YFDFWFEIFDMKKWDQFYSEVLGRPTTLSETMGKADVWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEDFVQSSGENGVVVFSLGSMVSNMTEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     IPLFADQPDNIAHMKARGAAVRVDFNTMSSTDLLNALKRVINDPSYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLVCVATVIFIV
                     TKCCLFCFWKFARKAKKGKND"
BASE COUNT      571 a    337 c    394 g    553 t
ORIGIN


  Query Match           84.6%;  Score 1450;  DB 93;  Length 1855;
  Best Local Similarity  91.7%;  Pred. No. 0;
  Matches 1568;  Conservative    0;  Mismatches  135;  Indels    7;  Gaps    3;

Qy      4 gcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccagtt 63
          |||||||||||||||| || |||||||||||||||  || |||| ||||| ||  || |
Db      2 GCATTGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTTTGCTAATACAACTG-AGCT 60

Qy     64 gttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccatt 123
            || |||||||||||||| ||||||||| |||||||||||||| ||||||||||||||||
Db     61 TTTGCTTTAGCTCTGGGAATTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATACAGCCATT 120
```

```
Qy    124 ggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtac 183
          ||||||||||| |||||||||||| | |||||| |||||||||||||||||||||||||||
Db    121 GGATGAATATAAAGACAATCCTGGATGAGCTTATTCAGAGAGGTCATGAGGTGACTGTAC 180

Qy    184 tggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaag 243
          |||||||||||||||||||||||||||||||  | |||||  |||||||| |||||
Db    181 TGGCATCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAA 240

Qy    244 tttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaaga 303
          ||||||| |||||||||||||||||||| || |||||| |||||||||||||| ||||||
Db    241 TTTATCCCACATCTTTAACTAAAACTGAGTTGGAGAATTTCATCATGCAACAGATTAAGA 300

Qy    304 gatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcc 363
          |||||||||| ||| |||||||  |||||||||||||||||||||| ||| ||||||
Db    301 GATGGTCAGACCTTCCAAAAGATACATTTTGGTTATATTTTTCACAAGTACAGGAAATCA 360

Qy    364 tgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaag 423
          |||   |||  |||||||  |||||| ||||||||||||||||||||||||||||||
Db    361 TGTCAATATTTGGTGACATAACTAGAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAAAT 420

Qy    424 ttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccct 483
          |||||||||| |||||||||||||||||||| || ||||||||||||| |||||||||
Db    421 TTATGAAAAAAGTACAAGAGTCAAGATTTGACGTCATTTTTGCAGATGCTATTTTTCCCT 480

Qy    484 gtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactc 543
          || |||||||||||||| |||| |||||||||  |||||||||||||| ||||  |||
Db    481 GTAGTGAGCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCTCAGCTTCTCTC 540

Qy    544 ctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacataccta 603
          ||||||||||  |||||| ||| |||||||||  | ||||||||||||||||||  |||||
Db    541 CTGGCTACACTTTTGAAAAGCATAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTG 600

Qy    604 ttgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatct 663
          |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGTTATGTCAGAATTAACTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCT 660

Qy    664 atgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagtttt 723
          |||||||||| |||||||||||||| |||| | ||| ||||||||||||||||||||||
Db    661 ATGTGCTTTACTTTGACTTTTGGTTCGAAATATTTGACATGAAGAAGTGGGATCAGTTTT 720

Qy    724 acagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatat 783
          | |||||||||| |||||||||||||||| |||| |||||||||||| |||||||||| |||
Db    721 ATAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTAT 780

Qy    784 ggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgatt 843
          ||||||| ||||||||||||| |||||| |||||||||||| |||||||||| |||||||
Db    781 GGCTTATTCGAAACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATT 840

Qy    844 ttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggagttt 903
          |||||||||| |||||| |||||||||||||||||| |||||||||||||| || |||
Db    841 TTGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTT 899

Qy    904 gtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaac 963
          |||||||||||||||||||||||||||||||||||||||||||||||||| || | ||||||
Db    900 GTACAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAAC 959

Qy    964 atgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggtt 1023
          ||||||| |||||||||||| ||||||||| ||||||| ||| ||||||||||||||||||
Db    960 ATGACAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTT 1019

Qy   1024 ctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataag 1083
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1020 CTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAG 1079

Qy   1084 tggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggt 1143
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   1080 TGGATACCCCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGT 1139

Qy   1144 ggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattg 1203
          |||||||||||||||| |||||||||||||||||||||||||||||||||| |||||||||
Db   1140 GGAGCCAATGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTG 1199

Qy   1204 ttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattg 1263
          |||   |||||||||||||||||||||||||||||||||||| |||||||||||||| ||
Db   1200 TTTGCCGATCAACCTGATAACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTG 1259

Qy   1264 gacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgat 1323
          |||||||||||||||||||||||||||| |||||||||| |||||| ||||||||||||
Db   1260 GACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGAT 1319

Qy   1324 cctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaag 1383
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db   1320 CCTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAG 1379

Qy   1384 cccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacac 1443
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   1380 CCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACAC 1439

Qy   1444 cttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattggg 1503
          ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1440 CTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGG 1499

Qy   1504 tttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgt 1563
          || ||||||| |||||||||||||||||||||||||| ||||||| | ||||||||||||
Db   1500 TTCCTGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGT 1559

Qy   1564 ttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacattt 1623
          |||||||||||||||||||  |||||||||||  |||||||||| ||||| ||||
Db   1560 TTCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTATATCTGAGATTT 1619

Qy   1624 gaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagat 1683
          |||||||||||||| ||||| | ||| ||||||||||||||||||||     ||||||
Db   1620 GAAGCTGGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAAG-----AAAGAT 1674

Qy   1684 tgttatgcaagatttcttttcttcctgtgac 1713
          ||| |||||||||||||||||||| |||
Db   1675 TGTGATGCAAGATTTCTTTCTTCCTGAGAC 1704
```

```
RESULT   4
AF016310
LOCUS       AF016310     1753 bp    mRNA            PRI       30-DEC-1998
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase (UGT2B18) mRNA,
            complete cds.
ACCESSION   AF016310
VERSION     AF016310.1  GI:4079706
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1753)
  AUTHORS   Beaulieu,M., Levesque,E., Barbier,O., Turgeon,D., Belanger,G.,
            Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Molecular Endocrinology, Laval University,
            2705 Laurier Blvd, Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .1753
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     gene            1. .1753
                     /gene="UGT2B18"
     CDS             1. .1590
                     /gene="UGT2B18"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAC98726.1"
                     /db_xref="GI:4079707"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENIIRQQIKRWSELPK
                     DTFWLYFSQMQEIMWKFGDITRNFCKDVVSNKKLMKKLQKSRFDVVFADAIFPCSELL
                     AELLNTPLVYSLRFTPGYNPEKHCGGFLFPPSYVPVVMSELSDHMTFMERVKNMIYML
                     YFDFCFQIYAMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
```

```
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVTNMKEERANVIASALAQIPQK
                     VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGSNGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKAKGAAVRLDFDTMSSTDLVNALKTVINDPLYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     MKCCLFCFWKFARKGKKGKSD"
BASE COUNT      549 a    323 c    362 g    519 t
ORIGIN


  Query Match             83.2%;  Score 1425.8;  DB 88;  Length 1753;
  Best Local Similarity   91.2%;  Pred. No. 0;
  Matches 1548;  Conservative    0;  Mismatches  142;  Indels    7;  Gaps    3;

Qy      17 atgactctgaaatggacttcagttcttctgctgatacatctccagttgttactttagctc 76
           ||| || |||||||||||||||  ||||||| ||||| ||  || | ||||||||||||
Db       1 ATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACT-GAGCTTTTACTTTAGCTC 59

Qy      77 tgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaa 136
           |||||||||||||| |||||||||||||| |||||||| ||||||||||||||||||||
Db      60 TGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCATTGGATGAATATGAA 119

Qy     137 gacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagc 196
           |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db     120 GACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTACTGGCATCTTCAGC 179

Qy     197 ttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatc 256
           |||||||||||||||||||  | ||||| |||||||| ||||||||| |||||||||
Db     180 TTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAGTTTTTCCTACATC 239

Qy     257 tttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacat 316
           |||||||||||||| |||||||||||||||| ||||||  |||||||||||||||  |
Db     240 TTTAACTAAAACTGAGTTTGAGAATATCATCAGGCAACAAATTAAGAGATGGTCAGAACT 299

Qy     317 tcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgggaattata 376
           || |||||||| |||||||||||||||||||   ||||||||| ||||| | ||
Db     300 TCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAATCATGTGGAAGTTTGG 359

Qy     377 tgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaact 436
           |||||||  |||||| ||||||||||||||||||||||||||||| | ||||| |||||
Db     360 TGACATAACTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGAAATTAATGAAGAAACT 419

Qy     437 acaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgct 496
           |||| |||||||||||||| |||||||||||||||| |||||||||||| |||| ||||
Db     420 ACAAAAGTCAAGATTTGACGTCGTTTTTGCAGATGCTATTTTTCCCTGTAGTGAGTTGCT 479

Qy     497 ggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggctacacaat 556
           |||||  |||||||||||  ||   |||||||||||||||||| ||||||||||||   |
Db     480 GGCTGAGCTACTTAACACACCCCTTGTGTACAGTCTCCGCTTCACTCCTGGCTACAATTT 539

Qy     557 tgaaaggcacagtggaggactgattttccctccttcctacatacctattgttatgtcaaa 616
           ||||| |||  |||||||| |  |||||||||| |||||  ||||| |||||||||||| |
Db     540 TGAAAAGCATTGTGGAGGATTTCTTTTCCCTCCCTCCTATGTACCTGTTGTTATGTCAGA 599

Qy     617 attaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtgctttattt 676
           |||||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     600 ATTAAGTGATCACATGACTTTCATGGAGAGGGTAAAAAATATGATCTATATGCTTTATTT 659

Qy     677 tgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagtgaagtttt 736
           ||||||||| |||||||| | ||  |||||||||||||||||||||| ||||||||| |
Db     660 TGACTTTTGTTTCCAAATATATGCCATGAAGAAGTGGGATCAGTTTTATAGTGAAGTTCT 719

Qy     737 aggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaa 796
           |||||||||||||| |||| |||||||||||| |||||||||||||||||||| |||||
Db     720 AGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACATATGGCTTATTCGAAA 779

Qy     797 ctcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggatt 856
           |||||||| |||||| ||||| |||||| |||||||||| |||||||||||||||||| |
Db     780 CTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGATTTTGTTGGAGGACT 839

Qy     857 ccactggcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacagagctctg 916
           ||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     840 CCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTTGTACAGAGCTCTG 898

Qy     917 gagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaa 976
           |||||||||||||||||||||||||||||||||||||||||| || | | |||||||| || |||||
Db     899 GAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCACTAACATGAAAGAAGAAA 958

Qy     977 gggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttg 1036
           ||||||| |||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db     959 GGGCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTG 1018

Qy    1037 atgggaataaaccagatgccttaggtctcaatactcggctgtataagtggataccccaga 1096
           |||| || |||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db    1019 ATGGAAAAAAACCAGATACCTTAGGTCTCAATACTCGGCTCTATAAGTGGATACCCCAGA 1078

Qy    1097 atgaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggca 1156
           |||| |||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||||
Db    1079 ATGATCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGATCCAATGGCA 1138

Qy    1157 tctatgaggcaatctaccatgggatccctatggtgggcattccattgttttgggatcaac 1216
           |||| |||||||||||||||| |||||||||||||||||||||||||||   |||||||
Db    1139 TCTACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCATTGTTTGCCGATCAAC 1198

Qy    1217 ctgataacattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaa 1276
           |||||||||||||||||||||| ||||| ||||||||||||||||||||||| ||||||
Db    1199 CTGATAACATTGCTCACATGAAAGCCAAAGGAGCAGCTGTTAGATTGGACTTCGACACAA 1258

Qy    1277 tgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttttatataaag 1336
           |||||||||||||| || ||||||| |||||||||||||||||||||||||||||||||||
Db    1259 TGTCGAGTACAGACTTGGTGAATGCCCTGAAGACAGTAATTAATGATCCTTTATATAAAG 1318

Qy    1337 agaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgag 1396
           ||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1319 AGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTGGATCGAG 1378

Qy    1397 cagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcag 1456
           |||||||||||||||||||||||||||||| |||||||||||| ||||||||||   |||||
Db    1379 CAGTCTTCTGGATTGAATTTGTCATGCGCCCACAAAGGAGCTAAACACCTTCGGCCTGCAG 1438

Qy    1457 cccatgacctcacctggttccagtaccactctttggatgtgattgggtttctgctggcct 1516
           |||||||||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db    1439 CCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTGGGTTCCTGCTGGCCT 1498

Qy    1517 gtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctggaagtttg 1576
           |||||||||||||||||||||||||||  || | ||||||||||||||||||||||||||
Db    1499 GTGTGGCAACTGTGATATTTATCATCATGAAATGTTGTCTGTTTTGTTTCTGGAAGTTTG 1558

Qy    1577 ctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagctggaaaac 1636
           |||||||||| |||||||||||||| |||||||||| ||||| ||||||||||||||||||
Db    1559 CTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTGAGATTTGAAGCTGGAAAAC 1618

Qy    1637 cagatagatagga caacttcagtttattccagcaagaaagaaaagattgttatgcaagat 1696
           | |||||||  ||| |||||||||||||||||||||||||||||     |||| || | || ||||||
Db    1619 CTGATAGATGAGACTACTTCAGTTTATTCCAGCAAG-----AAAGGTTATGATACAAGAT 1673

Qy    1697 ttctttcttcctgtgac 1713
           |||||||||||||||||
Db    1674 TTCTTTCTTCCTGTGAC 1690


RESULT   5
MFU91582
LOCUS       MFU91582     2648 bp    mRNA            PRI       25-MAR-1997
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION   U91582
VERSION     U91582.1  GI:1906053
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2648)
  AUTHORS   Belanger,G., Beaulieu,M., Levesque,E., Hum,W.H. and Belanger,A.
  TITLE     Expression and characterization of a novel
            UDP-Glucuronosyltransferase, UGT2B9, from cynomolgus monkey
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2648)
  AUTHORS   Belanger,G., Beaulieu,M., Levesque,E., Hum,W.H. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1997) Endocrinology, CHUL, 2705 Boulevard
            Laurier, Ste Foy, QC G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2648
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             28. .1617
                     /note="UGT2B9"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAB50249.1"
                     /db_xref="GI:1906054"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENISMQEVKRWIELPK
                     DTFWLYFSQMQEIMWRFGDIIRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYIFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     SFDFYFQMYDMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVTNMEEERANVIASALAQIPQK
                     VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTKGAAVRLDFDTMSSTDLANRLKTVINDPLYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFVI
                     MKCCLFCFWKFARKGKKGKSD"
BASE COUNT      821 a    499 c    562 g    766 t
ORIGIN
```

```
  Query Match           82.9%;  Score 1419.4;  DB 93;  Length 2648;
  Best Local Similarity   90.5%;  Pred. No. 0;
  Matches 1550;  Conservative    0;  Mismatches  156;  Indels    7;  Gaps    3;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
           || ||||||||||||||||  || ||||||||||||||||  |||||||  |||||  ||  |
Db      12 ATTGCATTGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACT-GA 70

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           | | ||||||||||||||||||||||||||||| |||||||||||||| ||||||||  ||||
Db      71 GCTTTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCC 130

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           ||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db     131 ATTGGATGAATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTG 190

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaaatttg 240
           | ||||||||||||||||||||||||||||||||||  | |||||  |||||||| |||
Db     191 TCCTGGCATCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTG 250

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||| | ||||||||||||||||||||||| ||||||||||||  |||||| ||||||
Db     251 AAGTCTTTCCTACATCTTTAACTAAAACTGAGTTTGAGAATATCTCCATGCAAGAGGTTA 310

Qy     301 agagatggtcagacattcgaaaagatagctttggttatatttttcacaagaacaagaaa 360
           ||||||||  |||  ||| ||||||||  ||||||||||||||||||||   |||||||
Db     311 AGAGATGGATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAA 370

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
           || |||||   ||   ||||||| ||||||| ||||||||||||||||||||||||||||
Db     371 TCATGTGGAGGTTTGGTGACATAATTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGA 430

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
           || | ||||| ||| ||||||||||||||||||| ||||||||||||| || |||||||
Db     431 AATTAATGAAGAAATTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTC 490

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
           ||||| |||||||||||||| |||| |||||||||   ||||||||||||||||||||| |
Db     491 CCTGTAGTGAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCA 550

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
           ||||||||||||   |||||| |||  |||||||  | |||||||||| |||||| |||
Db     551 CTCCTGGCTACATTTTTGAAAAGCATTGTGGAGGTTTTCTTTTCCCTCCCTCCTACGTAC 610

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
           || ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     611 CTGTTGTTATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGA 670

Qy     661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
           ||||| |||||| |||||||||||  || ||||||| ||| |||||||||||||||||| |
Db     671 TCTATATGCTTTCTTTTGACTTTTATTTTCAAATGTATGACATGAAGAAGTGGGATCAAT 730

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
           |||| |||||||||| |||||||||||||||| |||| |||||||||||| ||||||||||
Db     731 TTTATAGTGAAGTTCTAGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACA 790

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
           |||||||||| |||||||||||| |||||| ||||| |||||| |||||||||| ||||
Db     791 TATGGCTTATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTG 850

Qy     841 attttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggag 900
           |||||||||||||| |||||| ||||||||||||||||||||||||||||||||| |||
Db     851 ATTTTGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAG 909

Qy     901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| || | | |
Db     910 TTTGTACAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCACT 969

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
           ||||||  || |||||||||||| |||||||| |||||||||| ||||||||||||||
Db     970 AACATGGAAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAG 1029

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
           |||||||||||||||||||| || ||||| ||| ||||||| |||||||||||||| |||
Db    1030 GTTCTGTGGAGATTTGATGGAAAAAAACCGGATACCTTAGGCCTCAATACTCGGCTCTAT 1089

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttataactcat 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090 AAGTGGATACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCAC 1149

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           |||||||||||||||||||||| || |||||||||||| ||||||||||||||||||||
Db    1150 GGTGGAGCCAATGGCATCTATGAAGCGATCTACCATGGGGTCCCTATGGTGGGCATTCCA 1209

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
           ||||||   |||||||||||||||||||||||||| |||| ||||||||||||||||||
Db    1210 TTGTTTGCCGATCAGCCTGATAACATTGCTCACATGAAGACCAAAGGAGCAGCTGTTAGA 1269

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
           |||||||| |||||||||||||||||||||||| ||  |||| ||||||||||||||||||
Db    1270 TTGGACTTCGACACAATGTCGAGTACAGACTTGGCGAATCGACTGAAGACAGTAATTAAT 1329

Qy    1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1330 GATCCTTTATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTG 1389

Qy    1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db    1390 AAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAA 1449

Qy    1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1450 CACCTTCGGCCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATT 1509

Qy    1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
           ||||| |||||||||||||||||||||||||||| | |||| || | ||||||||||
Db    1510 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTTGTTATCATGAAATGTTGTCTGTTT 1569

Qy    1561 tgtttctggaagtttgctagaaaaggaagaagggaaaaagagattagttatgtctgaca 1620
           |||||||||||||||||||||||||| |||||||||||||| |||||||||| ||||| |
Db    1570 TGTTTCTGGAAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTGAGA 1629

Qy    1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
           |||||||||||||||| |||||||  ||| |||||||  |||||||||  |||     |||
Db    1630 TTTGAAGCTGGAAAACCTGATAGATGAGACTACTTCAGCTTATTCCAGGAAG-----AAA 1684
```

```
Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
          | || | ||||||||||||||||||||||||||||
Db   1685 GGTTATGATGCAAGATTTCTTTCTTCCTGTGAC 1717


RESULT   6
AF112113
LOCUS       AF112113     1768 bp    mRNA            PRI       02-DEC-1999
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase 2B23 precursor,
            mRNA, complete cds.
ACCESSION   AF112113
VERSION     AF112113.1  GI:6502548
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1768)
  AUTHORS   Barbier,O., Levesque,E., Belanger,A. and Hum,D.W.
  TITLE     UGT2B23, a novel uridine diphosphate-glucuronosyltransferase enzyme
            expressed in steroid target tissues that conjugates androgen and
            estrogen metabolites
  JOURNAL   Endocrinology 140 (12), 5538-5548 (1999)
  MEDLINE   20043918
   PUBMED   10579317
REFERENCE   2  (bases 1 to 1768)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .1768
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             28. .1617
                     /EC_number="2.4.1.17"
                     /note="UGT2B23"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B23 precursor"
                     /protein_id="AAF14353.1"
                     /db_xref="GI:6502549"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTALASSASILFDPNNSSALKIEVFPTSLPKPEFENIVTQEIKRWIELPK
                     DTFWLYFSQMQEIMWKFGDIFRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYVFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     YFDFCFQIYDMKKWDQFYTEVLGRHTTLSEIMGKADIWLIRNSWNFQFPHPLLPNVDF
                     IGGLLCKPAKPLPKEMEEFVQSSGENGVVVFTLGSMITNMKEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGVNTRLYKWIPQNDLLGHPKTKAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTRGAAVQLDFDTMSSTDLVNALKTVINDPLYKENVMKLSRIQR
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSFDVIGFLLACVATVIFII
                     MKCCLFCFWKFARKGKKGKSD"
BASE COUNT      546 a    327 c    374 g    521 t
ORIGIN


  Query Match             82.4%;  Score 1411.4;  DB 88;  Length 1768;
  Best Local Similarity   90.2%;  Pred. No. 0;
  Matches 1545;  Conservative    0;  Mismatches  161;  Indels    7;  Gaps    3;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
             || |  |||||||||||||| || ||||||||||||||||  |||||| |||| ||  |
Db      12 ATTGTTTTGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACTG-A 70

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           | | ||||||||||||||||||||||||| |||||||||||||| |||||||| ||||
Db      71 GCTTTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCC 130

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||
Db     131 ATTGGATGAATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTG 190

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           |||||||||||||||||||||||||||||||||||||||||  | ||||| |||||||| |||
Db     191 CACTGGCATCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTG 250

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||||| |||||||||||||| ||||| |||| |||||||||||  ||| |||| || |||
Db     251 AAGTTTTTCCTACATCTTTACCTAAACCTGAGTTTGAGAATATAGTCACGCAAGAGATTA 310

Qy     301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
           |||||||| ||| ||| |||||||| |||||||||||||||||||||||  |||||||
Db     311 AGAGATGGATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAA 370

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
           || ||||| | ||   |||||||||||||||||||||||||||||||||||||||||||
Db     371 TCATGTGGAAGTTTGGTGACATATTTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGA 430

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttttttc 480
           || | ||||| |||||||||||||||||||||||| ||||||||||||||| || ||||||
Db     431 AATTAATGAAGAAACTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTC 490

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
           ||||| |||||||||||||| |||| |||||||||   |||||||||||||||||||| |
Db     491 CCTGTAGTGAGCTGCTGGCTGAGCTATTTAACATACCACTTGTGTACAGTCTCCGCTTCA 550

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
           ||||||||||     |||||| ||| |||||||| |  |||||||||| |||||| |||
Db     551 CTCCTGGCTATGTTTTTGAAAAGCATTGTGGAGGATTTCTTTTCCCTCCCTCCTACGTAC 610

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
           || |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     611 CTGTTGTTATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGA 670

Qy     661 tctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
           ||||| |||||||||||||||||| |||||||| | ||| ||||||||||||||||||||
Db     671 TCTATATGCTTTATTTTGACTTTTGTTTCCAAATATATGACATGAAGAAGTGGGATCAGT 730

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
           |||| | ||||||| |||||||| |||||| |||| ||||| |||||| ||||||||||
Db     731 TTTATACTGAAGTTCTAGGAAGACACACTACATTATCTGAGATAATGGGGAAAGCTGACA 790

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
           |||||||||| |||||||||||| |||||| |||||| |||||||||| ||||
Db     791 TATGGCTTATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTG 850

Qy     841 attttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggag 900
           ||||| |||||||| ||| || |||||||||||||||||||||||||||||||||| |||
Db     851 ATTTTATTGGAGGACTCCTCT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAG 909

Qy     901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||| | | |
Db     910 TTTGTACAGAGCTCTGGAGAAAACGGTGTTGTGGTGTTTACTCTGGGGTCAATGATCACT 969

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
           ||||||| || |||||||||||| |||||||| |||||||||| |||| |||||||||
Db     970 AACATGAAAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATTCCACAAAAG 1029

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
           |||||||||||||||||||||||||||||||||| |||||||| |||||||||||| |||
Db    1030 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTGTCAATACTCGGCTCTAT 1089

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1090 AAGTGGATACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAAAGCTTTTATAACTCAC 1149

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db    1150 GGTGGAGCCAATGGCATCTACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCA 1209

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
           |||||  |||||||||||||||||||||||||||||||  |||  ||||||||||||  |
Db    1210 CTGTTTGCCGATCAACCTGATAACATTGCTCACATGAAAACCAGAGGAGCAGCTGTTCAA 1269

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
           ||||||||| |||||||||||||||||||| || ||||||||||||||||||||||||||
```

```
Db   1270 TTGGACTTCGACACAATGTCGAGTACAGACTTGGTGAATGCACTGAAGACAGTAATTAAT 1329

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
          |||||||||||||||||||| |||||||||||||||||||||||| |||||| |||||
Db   1330 GATCCTTTATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACGTGATCAGCCAGTG 1389

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db   1390 AAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAA 1449

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
          ||||||||  |||||||||||||||||||||||||||||| |||||||| |||||||||
Db   1450 CACCTTCGGCCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTCGATGTGATT 1509

Qy   1501 gggttttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
          |||||  |||||||||||||||||  |||||||||||||||||||  || | ||||||||||
Db   1510 GGGTTCCTGCTGGCCTGTGTGGCCACTGTGATATTTATCATCATGAAATGTTGTCTGTTT 1569

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
          |||||||||||||||||||||||||| |||||||||||||| |||||||| | ||||| |
Db   1570 TGTTTCTGGAAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTGTATCTGAGA 1629

Qy   1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
          |||||||||||||||| |||||||  ||| |||||||||||||||| |||     |||
Db   1630 TTTGAAGCTGGAAAACCTGATAGATGAGACTACTTCAGTTTATTCCAGGAAG-----AAA 1684

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
          | || | |||||||||||||||||||||||||||
Db   1685 GGTTATGATGCAAGATTTCTTTCTTCCTGTGAC 1717
```

```
RESULT   7
AF294901
LOCUS       AF294901     1662 bp    mRNA            PRI       05-OCT-2000
DEFINITION  Macaca mulatta UDP-Glucuronosyltransferase UGT2B9*2 mRNA, complete
            cds.
ACCESSION   AF294901
VERSION     AF294901.1  GI:10644704
KEYWORDS    .
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Nucleotide and Deduced Amino Acid Sequence of UGT2B9*2: a Novel
            UDP-Glucuronosyltransferase Isoform Cloned from Female Rhesus
            Monkey Liver
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2000) Drug Metabolism, Merck & Co., Mail Drop
            RY80L-109, PO Box 2000, Rahway, NJ 07065, USA
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /sex="female"
                     /tissue_type="liver"
     CDS             43. .1632
                     /note="UGT"
                     /codon_start=1
                     /product="UDP-Glucuronosyltransferase UGT2B9*2"
                     /protein_id="AAG21377.1"
                     /db_xref="GI:10644705"
                     /translation="MSVKWTSVILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENISMQEVKRWIELPK
                     DTFWLYFSQMQEIMWRFGDIIRNFCKDVVSNKKLMKKLQESRFDVVFADPIFPCSELL
                     AELFNIPLVYSLRFTPGYIFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYML
                     SFDFYFQMYDMKKWDQFYSEVLGRPTTLSETMGKADIWLIRNSWNFQFPHPLLPNVDF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMVTNMEEERANVIASALAQIPQK
                     VLWRFDGKKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTKGAAVRLDFDTMSSTDLANALKTVINDPLYKENVMKLSRIQH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVILVI
                     MKCCLLCFWKFARKGKKGKSD"
BASE COUNT      479 a    319 c    367 g    497 t
ORIGIN
```

```
  Query Match             79.1%;  Score 1355.2;  DB 88;  Length 1662;
  Best Local Similarity   90.7%;  Pred. No. 0;
  Matches 1466;  Conservative    0;  Mismatches  148;  Indels    2;  Gaps    2;

Qy       2 tcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccag 61
           | |||||||||||||||| || ||||||||||||||||  |||||||| ||||| ||  ||
Db      28 TTGCATTGCACCAGGATGTCTGTGAAATGGACTTCAGTAATTCTGCTAATACAACT-GAG 86

Qy      62 ttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcca 121
           | |||||||||||||||||||||||||||| |||||||||||||| |||||||| |||||
Db      87 CTTTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCA 146

Qy     122 ttggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgt 181
           |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db     147 TTGGATGAATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGAGGTCATGAGGTGACTGT 206

Qy     182 actggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttga 241
           |||||||||||||||||||||||||||||||||||||  | ||||| |||||||| ||||
Db     207 CCTGGCATCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGA 266

Qy     242 agtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaa 301
           ||| | |||||||||||||||||||||||| ||||||||||  ||||||| |||||||
Db     267 AGTCTTTCCTACATCTTTAACTAAAACTGAGTTTGAGAATATCTCCATGCAAGAGGTTAA 326

Qy     302 gagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaat 361
           |||||||  |||  ||| ||||||||  |||||||||||||||||||||   ||||||||
Db     327 GAGATGGATAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATGCAAGAAAT 386

Qy     362 cctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaa 421
           | |||||   ||   ||||||| |||||| ||||||||||||||||||||||||||||||
Db     387 CATGTGGAGGTTTGGTGACATAATTAGAAATTTCTGTAAAGATGTAGTTTCAAATAAGAA 446

Qy     422 agttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttcc 481
           | | ||||| ||| |||||||||||||||||||| |||||||||||||| || |||||||
Db     447 ATTAATGAAGAAATTACAAGAGTCAAGATTTGACGTCGTTTTTGCAGATCCTATTTTTCC 506

Qy     482 ctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttac 541
           |||| |||||||||||||| |||| |||||||||   |||||||||||||||||||| ||
Db     507 CTGTAGTGAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCAC 566

Qy     542 tcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacc 601
           ||||||||||   |||||| |||  |||||||  |  |||||||||  |||||| ||||
Db     567 TCCTGGCTACATTTTTGAAAAGCATTGTGGAGGTTTTCTTTTCCCTCCCTCCTACGTACC 626

Qy     602 tattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgat 661
           | |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     627 TGTTGTTATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGAT 686

Qy     662 ctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagtt 721
           |||| ||||||  |||||||||||  || |||||||  ||| ||||||||||||||||  ||
Db     687 CTATATGCTTTCTTTTGACTTTTATTTTCAAATGTATGACATGAAGAAGTGGGATCAATT 746

Qy     722 ttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacat 781
           ||| |||||||| |||||||||||||||| |||| ||||||||||| |||||||||||||
Db     747 TTATAGTGAAGTTCTAGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACAT 806

Qy     782 atggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttga 841
           |||||||| |||||||||||| |||||| ||||| |||||| |||||||||| |||||
Db     807 ATGGCTTATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGA 866

Qy     842 ttttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggagt 901
           |||||||||||||| |||||| ||||||||||||||||||||||||||||||||| ||||
Db     867 TTTTGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGT 925

Qy     902 ttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagta 961
```

```
Db    926 TTGTACAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCAATGGTCACTA 985

Qy    962 acatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaagg 1021

Db    986 ACATGGAAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATCCCACAAAAGG 1045

Qy   1022 ttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtata 1081

Db   1046 TTCTGTGGAGATTTGATGGGAAAAAACCGGATACCTTAGGCCTCAATACTCGGCTCTATA 1105

Qy   1082 agtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatg 1141

Db   1106 AGTGGATACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCACG 1165

Qy   1142 gtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccat 1201

Db   1166 GTGGAGCCAATGGCATCTACGAAGCGATCTACCATGGGGTCCCTATGGTGGGCATTCCAT 1225

Qy   1202 tgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagat 1261

Db   1226 TGTTTGCCGATCAGCCTGATAACATTGCTCACATGAAGACCAAAGGAGCAGCTGTTAGAT 1285

Qy   1262 tggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatg 1321

Db   1286 TGGACTTTGACACAATGTCGAGTACAGACTTGGCGAATGCACTGAAGACAGTAATTAATG 1345

Qy   1322 atcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaa 1381

Db   1346 ATCCTTTATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGA 1405

Qy   1382 agcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaac 1441

Db   1406 AGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAAC 1465

Qy   1442 accttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattg 1501

Db   1466 ACCTTCGGCCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTG 1525

Qy   1502 ggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgtttt 1561

Db   1526 GGTTCCTGCTGGCCTGTGTGGCAACTGTGATACTTGTTATCATGAAATGTTGTCTGCTTT 1585

Qy   1562 gtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1617

Db   1586 GTTTCTGGAAGTTTGCTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTG 1641
```

```
RESULT   8
HSAJ5162
LOCUS       HSAJ5162     2079 bp    mRNA            PRI       24-JUN-1998
DEFINITION  Homo sapiens mRNA for UDP-glucuronosyltransferase.
ACCESSION   AJ005162
VERSION     AJ005162.1  GI:3135024
KEYWORDS    UDP-glucuronosyltransferase; UGT2B4 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2079)
  AUTHORS   Ritter,J.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) Ritter J.K., Pharmacology and Toxicology,
            Virginia Commonwealth University, P.O. Box 980613, Medical College
            of Virginia, Richmond, Virginia, 23112-0613, USA
REFERENCE   2  (bases 1 to 2079)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Lubet,R.A. and Owens,I.S.
  TITLE     Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
            log differences in activity toward parallel substrates including
            hyodeoxycholic acid and certain estrogen derivatives
  JOURNAL   Biochemistry 31 (13), 3409-3414 (1992)
  MEDLINE   92207964
REFERENCE   3  (bases 1 to 2079)
  AUTHORS   Jackson,M.R., McCarthy,L.R., Harding,D., Wilson,S., Coughtrie,M.W.
            and Burchell,B.
  TITLE     Cloning of a human liver microsomal UDP-glucuronosyltransferase
            cDNA
  JOURNAL   Biochem. J. 242 (2), 581-588 (1987)
  MEDLINE   87241362
FEATURES             Location/Qualifiers
     source          1. .2079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UDPGTh-1"
                     /clone_lib="Lambda bacteriophage gt11"
                     /tissue_lib="Liver"
     gene            34. .1620
                     /gene="UGT2B4"
     CDS             34. .1620
                     /gene="UGT2B4"
                     /EC_number="2.4.1.17"
                     /function="transfers glucuronic acid to hyodeoxycholic
                     acid and estrogen catechols"
                     /codon_start=1
                     /evidence=experimental
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="CAA06396.1"
                     /db_xref="GI:3135025"
                     /db_xref="SPTREMBL:O60867"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
     conflict        546
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="g"
     conflict        548. .549
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="ct"
     conflict        1187. .1188
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="ct"
     conflict        1190
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="c"
     conflict        1192
                     /gene="UGT2B4"
                     /citation=[3]
                     /replace="a"
     conflict        1977. .1978
                     /citation=[3]
                     /replace="cga"
     conflict        2068
                     /citation=[3]
                     /replace="g"
BASE COUNT      634 a    396 c    436 g    613 t
ORIGIN
```

```
  Query Match           79.1%;  Score 1354.2;  DB 92;  Length 2079;
  Best Local Similarity  88.8%;  Pred. No. 0;
  Matches 1522;  Conservative    0;  Mismatches  183;  Indels    8;  Gaps    5;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60

Db     18 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCT-GA 76
```

```
Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           | |||||||||||||||||||||||||||||| |||||||||||| || |||||| |||||
Db      77 GCTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 136

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           | |||||||||||| |||||||||||| | || ||||| |||||||||||||||||||||
Db     137 ACTGGATGAATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTG 196

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           || |||||||||||||||||| ||| ||||||||     |||| ||||||||||||||
Db     197 TATTGGCATCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTG 256

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||||||||||  |||||||||||||||||| |||||| |||| |||| ||| | |||||
Db     257 AAGTTTATCCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTA 316

Qy     301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
           |||||||| ||||  ||| |||||| |  ||||||| |||||||||||||| ||||||||
Db     317 AGAGATGGGCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAA 376

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
           || |||||  |||  |||||||| ||||||| |||||||| ||| ||||||||||||||||
Db     377 TCATGTGGACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGA 436

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
           || ||||||| |||||||| |||||||||||||  | ||| |||||||||||||||| |
Db     437 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCC 496

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
           ||| |||||||||||||| | | ||||||| ||||  ||||| ||||| ||||||||
Db     497 CCTTTGGTGAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCT 556

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
           |||||||||| |||||||| ||| |||||||||||  | ||||||||||||||  | |
Db     557 CTCCTGGCTACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGC 616

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
           || |||||||||| || ||||||| |||||||||||||| ||||||||||||||||||||
Db     617 CTGTTGTTATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGA 676

Qy     661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
           |||||||||||||||||| |||||||||||||| | ||| ||||||||||||||||||||
Db     677 TCTATGTGCTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGT 736

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
           | |||||||||||| |||||||||||||||| |||| |||||||||| ||||||||||
Db     737 TCTACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACA 796

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
           |||||||||| ||||||| |||| ||||||||||||||| ||| |||||||||| ||||
Db     797 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTG 856

Qy     841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
           | || |||||||||| |||||| |||||||||||||||||||| |||||||||| |||
Db     857 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAG 915

Qy     901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
           ||||| |||||||||||||||||||||||||||||||||||||||||||  || | |||
Db     916 TTTGTCCAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGT 975

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
           |||| | ||| |||||||||||||||||||| ||||||||||||||||||||||||||
Db     976 AACACGTCAGAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1035

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
           |||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||
Db    1036 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTAC 1095

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
           |||||||||||||||||| ||||| ||||| ||||||||||||||||||||||||||||
Db    1096 AAGTGGATACCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCAT 1155

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1156 GGTGGAGCCAATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCA 1215

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
           ||||||   |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1216 TTGTTTGCAGATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGT 1275

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
           ||||||||| |||||||||||||||||||| | || ||||||||||||||||||||||||
Db    1276 TTGGACTTCCACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAAT 1335

Qy    1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
           |||||||||||||||||||  ||||||||||||||||||| ||||||||||||||||||
Db    1336 GATCCTTTATATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTG 1395

Qy    1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
           |||||||| |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db    1396 AAGCCCCTTGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAG 1455

Qy    1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
           |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||| |
Db    1456 CACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACT 1515

Qy    1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
           ||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1516 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTT 1572

Qy    1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
           ||| |||||||||||| ||||| ||| |||||||| |||||||||| ||| ||||||
Db    1573 TGTGTCTGGAAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGG 1632

Qy    1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
           | |||||||| |||||  ||| || | ||  ||| |||||||| || ||||||   |
Db    1633 CTGGAAGCTGGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACG 1689

Qy    1681 gattgttatgcaagatttctttcttcctgtgac 1713
              || ||    ||||| |||||||| ||||||
Db    1690 TTGTGATACAAGAGATTCCTTTCTTCTTGTGAC 1722
```

```
RESULT   9
AF294902
LOCUS       AF294902     1662 bp    mRNA            PRI       05-OCT-2000
DEFINITION  Macaca mulatta UDP-Glucuronosyltransferase UGT2B33 mRNA, complete
            cds.
ACCESSION   AF294902
VERSION     AF294902.1  GI:10644706
KEYWORDS    .
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Nucleotide and Deduced Amino Acid Sequence of UGT2B33: a Novel
            UDP-Glucuronosyltransferase Isoform Cloned from Female Rhesus
            Monkey Liver
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1662)
  AUTHORS   Dean,B., Zhao,S. and King,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2000) Drug Metabolism, Merck & Co., Mail Drop
            RY80L-109, PO Box 2000, Rahway, NJ 07065, USA
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /sex="female"
                     /tissue_type="liver"
     CDS             43. .1632
                     /note="UGT"
```

```
                     /codon_start=1
                     /product="UDP-Glucuronosyltransferase UGT2B33"
                     /protein_id="AAG21378.1"
                     /db_xref="GI:10644707"
                     /translation="MSVKWTSIILLIQLSFYFSSGSCGKVLVWAAEYSHWMNMKTILE
                     ELVQRGHEVTVLASSASILFDPNNSSALKIEVFPTSLTKTEFENIIRQQIKRWSELPK
                     DTFWLYFSQIQEIMWRFGDISIKFCKDVVSNKKLMKKLQESRFDVVLADPIFPCSELL
                     AELFNIPLVYSLRFTPGYVFEKHCGGFLFPPSYVPVVMSELSDQMTFMERVKNMIYVL
                     YFDFCFQLYDMKKWDQFYSEVLGRHTTLSEIMGKADIWLIRNSWNFQFPHPLLPNVDF
                     IGGLLCKPAKPLPKEMEEFVQSSGENGVVVFTLGSMITNMKEERANVIASALAQIPQK
                     VLWRFDGNKPDTLGVNTRLYKWIPQNDLLGHPKTKAFITHGGANGIYEAIYHGVPMVG
                     IPLFADQPDNIAHMKTRGAAVQLDFDTMSSTDLANALKTVINDPLYKENVMKLSRIQR
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     MKCCLFCFWKFTRKGKKGKSD"
BASE COUNT      486 a    316 c    364 g    496 t
ORIGIN
```

Query Match 79.0%; Score 1353.6; DB 88; Length 1662;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1465; Conservative 0; Mismatches 149; Indels 2; Gaps 2;

```
Qy       2 tcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccag 61
          | |||||||||||||||| || |||||||||||||| | |||||| ||||| ||  ||
Db      28 TTGCATTGCACCAGGATGTCTGTGAAATGGACTTCAATAATTCTGCTAATACAACTG-AG 86

Qy      62 ttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcca 121
           | |||||||||||||||||||||||||||||| |||||||||||| |||||||| |||||
Db      87 CTTTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATATAGCCA 146

Qy     122 ttggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgt 181
           |||||||||||||||||||||||| |||||||||| ||||| ||||||||||||||||||
Db     147 TTGGATGAATATGAAGACAATCCTGGAAGAGCTTGTCCAGAGGGGTCATGAGGTGACTGT 206

Qy     182 actggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttga 241
           ||||||||||||||||||||||||||||||||||  | ||||| |||||||| ||||
Db     207 ACTGGCATCTTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGA 266

Qy     242 agtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaa 301
           ||||| |||||||||||||||||||| || |||||||||||||| |||||| ||||
Db     267 AGTTTTTCCTACATCTTTAACTAAAACGGAGTTTGAGAATATCATCAGGCAACAGATTAA 326

Qy     302 gagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaat 361
           |||||||||||  ||| ||||||||  ||||||||||||||||||||  |||||||||
Db     327 GAGATGGTCAGAACTTCCAAAAGATACATTTTGGTTATATTTTTCACAAATACAAGAAAT 386

Qy     362 cctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaa 421
           | |||||   ||   |||||||  || ||| ||||||||||||||||||||||||||||||
Db     387 CATGTGGAGGTTTGGTGACATAAGTATAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAA 446

Qy     422 agttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttcc 481
           | | ||||| |||||||||||||||||||||||  |||||||  |||||| || |||||||
Db     447 ATTAATGAAGAAACTACAAGAGTCAAGATTTGATGTCGTTTTAGCAGATCCTATTTTTCC 506

Qy     482 ctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttac 541
           |||| ||||||||||||||| |||| |||||||||   ||||||||||||||||||| ||
Db     507 CTGTAGTGAGCTGCTGGCTGAGCTATTTAACATACCCCTTGTGTACAGTCTCCGCTTCAC 566

Qy     542 tcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacc 601
           ||||||||||    ||||||  ||  |||||||| |  |||||||||| |||||| ||||
Db     567 TCCTGGCTACGTTTTTGAAAAACATTGTGGAGGATTTCTTTTCCCTCCCTCCTACGTACC 626

Qy     602 tattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgat 661
           | ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     627 TGTTGTTATGTCAGAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGAT 686

Qy     662 ctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatcagtt 721
           |||||||||||||||||||||||| |||||| ||| ||| ||||||||||||||||||||
Db     687 CTATGTGCTTTATTTTGACTTTTGTTTCCAATTGTATGACATGAAGAAGTGGGATCAGTT 746

Qy     722 ttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacat 781
           ||| ||||||||| ||||||||| |||||| |||| ||||| |||||| ||||||||||||
Db     747 TTATAGTGAAGTTCTAGGAAGACACACTACATTATCTGAGATAATGGGGAAAGCTGACAT 806
```

```
Qy     782 atggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttga 841
           ||||||||| |||||||||||| |||||| ||||| |||||| |||||||||| |||||
Db     807 ATGGCTTATTCGAAACTCCTGGAATTTTCAGTTTCCACATCCACTCTTACCAAATGTTGA 866

Qy     842 ttttgttggaggattccactggcaaacctgccaaaacccctacctaaggaaatggaggagt 901
           |||| |||||||| ||| || ||||||||||||||||||||||||||||||||| || |
Db     867 TTTTATTGGAGGACTCCTCT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAAT 925

Qy     902 ttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagta 961
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||| | ||
Db     926 TTGTACAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTACTCTGGGGTCAATGATCACTA 985

Qy     962 acatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaagg 1021
           |||||| || |||||||||||| |||||||| |||||||||| |||| ||||||||||
Db     986 ACATGAAAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCCAGATTCCACAAAAGG 1045

Qy    1022 ttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtata 1081
           |||||||||||||||||||||||||||||||||| |||||||| |||||||||||| ||||
Db    1046 TTCTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTGTCAATACTCGGCTCTATA 1105

Qy    1082 agtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatg 1141
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db    1106 AGTGGATACCCCAGAATGACCTTCTAGGTCATCCAAAAACCAAAGCTTTTATAACTCACG 1165

Qy    1142 gtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccat 1201
           |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db    1166 GTGGAGCCAATGGCATCTACGAGGCAATCTACCATGGGGTCCCTATGGTGGGCATTCCAC 1225

Qy    1202 tgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagat 1261
           |||||  |||||||||||||||||||||||||||||  |||  |||||||||||  ||
Db    1226 TGTTTGCCGATCAACCTGATAACATTGCTCACATGAAAACCAGAGGAGCAGCTGTTCAAT 1285

Qy    1262 tggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatg 1321
           |||||||| ||||||||||||||||||||||||||| ||  ||||||||||||||||||||
Db    1286 TGGACTTCGACACAATGTCGAGTACAGACTTGGCGAATGCACTGAAGACAGTAATTAATG 1345

Qy    1322 atcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaa 1381
           |||||||||||||||||||| |||||||||||||||||||||||| |||||| ||||| |
Db    1346 ATCCTTTATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACGTGATCAGCCAGTGA 1405

Qy    1382 agcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaac 1441
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    1406 AGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAAC 1465

Qy    1442 accttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattg 1501
           |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1466 ACCTTCGGCCTGCAGCCCATGACCTCACCTGGTTCCAGTATCACTCTTTGGATGTGATTG 1525

Qy    1502 ggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgtttt 1561
           |||| |||||||||||||||| ||||||||||||||||||  || | ||||||||||
Db    1526 GGTTCCTGCTGGCCTGTGTGGCCACTGTGATATTTATCATCATGAAATGTTGTCTGTTTT 1585

Qy    1562 gtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1617
           |||||||||||||| |||||||||| |||||||||||||||| |||||||||| ||||
Db    1586 GTTTCTGGAAGTTTACTAGAAAAGGAAAGAAGGGAAAAAGTGATTAGTTATATCTG 1641
```

```
RESULT  10
AF064200
LOCUS       AF064200     2097 bp    mRNA            PRI       11-DEC-1998
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 2B4 precursor (UGT2B4)
            mRNA, UGT2B4*E458 allele, complete cds.
ACCESSION   AF064200
VERSION     AF064200.1  GI:3153831
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2097)
  AUTHORS   Levesque,E., Beaulieu,M., Hum,M. and Belanger,A.
  TITLE     Characterization and substrate specificity of UGT2B4(E458): A
```

```
          UDP-glucuronosyltransferase encoded by a polymorphic gene
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 2097)
  AUTHORS   Levesque,E., Beaulieu,M. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Ste-Foy, PQ G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2097
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .2097
                     /gene="UGT2B4"
                     /allele="UGT2B4*E458"
     CDS             26. .1612
                     /gene="UGT2B4"
                     /note="E458 variant"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B4 precursor"
                     /protein_id="AAC95002.1"
                     /db_xref="GI:3153832"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLERAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
BASE COUNT      653 a    393 c    435 g    616 t
ORIGIN
```

Query Match 78.9%; Score 1351; DB 88; Length 2097;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 1520; Conservative 0; Mismatches 185; Indels 8; Gaps 5;

```
Qy       1 atcgcattgcaccaggatgacttctgaaatggacttcagttcttctgctgatacatctcca 60
           || |||||||| ||||||| || |||||||||||||||| |||||||||||||||| ||  |
Db      10 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCT-GA 68

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           | |||||||||||||||||||||||||||||| |||||||||||| || |||||| |||||
Db      69 GCTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 128

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           | |||||||||| |||||||||||| | || ||||| ||||||||||||||||||||||||||
Db     129 ACTGGATGAATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTG 188

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           || |||||||||||||||||||| ||| ||||||||     |||| ||||||||||||
Db     189 TATTGGCATCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTG 248

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||||||||||  |||||||||||||||||| |||||| |||| |||| ||| | |||||
Db     249 AAGTTTATCCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTA 308

Qy     301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
           |||||||| ||||  ||| |||||| |  |||||| |||||||||||||| ||||||||
Db     309 AGAGATGGGCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAA 368

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
           || |||||  |||  |||||||| ||||||| |||||||| ||| ||||||||||||||||
Db     369 TCATGTGGACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGA 428

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttttttc 480
           || ||||||| |||||||| ||||||||||||||  | ||| ||||||||||||||| |
Db     429 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCC 488

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
           ||| |||||||||||||| | | ||||||| |||| ||||| ||||| ||||||||
```

```
Db     489 CCTTTGGTGAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCT 548

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
           |||||||||| |||||||| ||| |||||||||| | ||||||||||||| | |
Db     549 CTCCTGGCTACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGC 608

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
           || |||||||||| || ||||||| |||||||||||||| ||||||||||||||||||||
Db     609 CTGTTGTTATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGA 668

Qy     661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
           |||||||||||||||||| |||||||||||||| | ||| ||||||||||||||||||||
Db     669 TCTATGTGCTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGT 728

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
           | |||||||||| |||||||||||||||| |||| |||||||||| ||||||||||||
Db     729 TCTACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACA 788

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
           |||||||||| ||||||| |||| ||||||||||||||| ||| |||||||||| ||||
Db     789 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTG 848

Qy     841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
           | || ||||||||| |||||| ||||||||||||||||||||||| |||||||||| |||
Db     849 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAG 907

Qy     901 tttgtacagagctctggagaaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
           ||||| |||||||||||||||||||||||||||||||||||||||||||| || | |||
Db     908 TTTGTCCAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGT 967

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
           |||| | ||| |||||||||||||||||||||||||||| |||||||||||||||||||
Db     968 AACACGTCAGAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1027

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
           |||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||
Db    1028 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTAC 1087

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttataactcat 1140
           |||||||||||||||||| ||||| ||||| ||||||||||||||||||||||||||||
Db    1088 AAGTGGATACCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCAT 1147

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    1148 GGTGGAGCCAATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCA 1207

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
           ||||||   |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1208 TTGTTTGCAGATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGT 1267

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
           |||||||| |||||||||||||||||| | || ||||||||||||||||||||||||||
Db    1268 TTGGACTTCCACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAAT 1327

Qy    1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
           |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db    1328 GATCCTTTATATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTG 1387

Qy    1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
           |||||||| ||  |||||||||||||||||||||||||||||| ||| ||||||||||||
Db    1388 AAGCCCCTTGAAAGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAG 1447

Qy    1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1448 CACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACT 1507

Qy    1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
           ||||| |||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db    1508 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTT 1564

Qy    1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
           ||| |||||||||||| ||||| ||| |||||||| |||||||||| ||| ||||||
Db    1565 TGTGTCTGGAAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGG 1624
```

```
Qy   1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
          | ||||||||| |||||  ||| || | ||  ||| |||||||| || ||||||   |
Db   1625 CTGGAAGCTGGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACG 1681

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
             || ||    ||||| |||||||| ||||||
Db   1682 TTGTGATACAAGAGATTCCTTTCTTCTTGTGAC 1714


RESULT  11
AF081793
LOCUS       AF081793     2091 bp    mRNA            PRI       18-AUG-1998
DEFINITION  Homo sapiens UDP glucuronosyltransferase 2B4 precursor (UGT2B4)
            mRNA, UGT2B4*L109, L396 variant allele, complete cds.
ACCESSION   AF081793
VERSION     AF081793.1  GI:3426331
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2091)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
  MEDLINE   93326164
REFERENCE   2  (bases 1 to 2091)
  AUTHORS   Mackenzie,P.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1998) Clinical Pharmacology, Flinders University
            of South Australia, Bedford Park, SA 5042, Australia
FEATURES             Location/Qualifiers
     source          1. .2091
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
     gene            1. .2091
                     /gene="UGT2B4"
                     /note="formerly UGT2B11"
                     /allele="UGT2B4*L109, L396 variant"
     CDS             36. .1622
                     /gene="UGT2B4"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 2B4 precursor"
                     /protein_id="AAC32272.1"
                     /db_xref="GI:3426332"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYLSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKKWKEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLLADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
BASE COUNT      638 a    401 c    437 g    615 t
ORIGIN


  Query Match           78.5%;  Score 1344.6;  DB 88;  Length 2091;
  Best Local Similarity   88.5%;  Pred. No. 0;
  Matches 1516;  Conservative    0;  Mismatches  189;  Indels    8;  Gaps    5;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
          || |||||||| ||||||| || |||||||||||||| |||||||||||||||| ||  |
Db     20 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCT-GA 78

Qy     61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
          | |||||||||||||||||||||||||||||| |||||||||||| || |||||| |||||
Db     79 GCTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 138

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
          | |||||||||| |||||||||| | || ||||| ||||||||||||||||||||||
Db    139 ACTGGATGAATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTG 198

Qy    181 tactggcatcttcagcttccatttttttgatcccaatgatgcatccactcttaaatttg 240
          || |||||||||||||||||||  ||| ||||||||     |||| ||||||||||||
Db    199 TATTGGCATCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTG 258

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
          ||||||||||  |||||||||||||||||| |||||| |||| |||| ||| | ||||||
Db    259 AAGTTTATCCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTA 318

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
          |||||||| |||  ||| |||||| |  ||||||| |||| ||||||||| ||||||||
Db    319 AGAGATGGGCAGAACTTCCAAAAGACACATTTTGGTCATATCTTTCACAAGTACAAGAAA 378

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
          || |||||  |||  |||||||| ||||||| |||||||| ||| ||||||||||||||
Db    379 TCATGTGGACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGA 438

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
          || ||||||| |||||||| |||||||||||||  | ||| ||||||| |||||||| |
Db    439 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGACGCTGTTTTCC 498

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
          ||| |||||||||||||| | | ||||||| |||| ||||| ||||| ||||||||
Db    499 CCTTTGGTGAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCT 558

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
          |||||||||| |||||||| ||| ||||||||||  | |||||||||||||  | |
Db    559 CTCCTGGCTACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGC 618

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
          || |||||||||| || ||||||| |||||||||||||| ||||||||||||||||||
Db    619 CTGTTGTTATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGA 678

Qy    661 tctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
          |||||||||||||||||| |||||||||||||  | ||| ||||||||||||||||||
Db    679 TCTATGTGCTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGT 738

Qy    721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
          | |||||||||||| |||||||||||||||| |||| |||||||||| ||||||||||
Db    739 TCTACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACA 798

Qy    781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
          |||||||||| ||||||| ||||  |||||||||||||||| ||| ||||||||| ||||
Db    799 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTG 858

Qy    841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
          | || ||||||||| |||||| ||||||||||||||||||||||| ||| ||  | |||||
Db    859 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCGAAGAAATGGAAGGAG 917

Qy    901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
          ||||| ||||||||||||||||||||||||||||||||||||||||||  || | |||
Db    918 TTTGTCCAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGT 977

Qy    961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
          |||| | ||| |||||||||||||||||||||| ||||||||||||||||||||||||
Db    978 AACACGTCAGAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1037

Qy   1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
          |||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||
Db   1038 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTAC 1097

Qy   1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
          |||||||||||||||||| ||||| ||||| ||||||||||||||||||||||||||||
Db   1098 AAGTGGATACCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCAT 1157

Qy   1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1158 GGTGGAGCCAATGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCA 1217
```

```
Qy   1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
          ||| ||   |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1218 TTGCTTGCAGATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGT 1277

Qy   1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
          ||||||||| |||||||||||||||||||| | || ||||||||||||||||||||||||
Db   1278 TTGGACTTCCACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAAT 1337

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
          ||||||||||||||||||  |||||||||||||||||||| ||||||||||||||
Db   1338 GATCCTTTATATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTG 1397

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
          |||||||| |||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db   1398 AAGCCCCTTGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAG 1457

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
          |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||| |
Db   1458 CACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACT 1517

Qy   1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
          ||||| |||||||||||||||||||||||||||| ||||||||||    |||||||||
Db   1518 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTT 1574

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
          ||| |||||||||||| ||||| ||| |||||||| |||||||||| ||| ||||||
Db   1575 TGTGTCTGGAAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGG 1634

Qy   1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
           | |||||||| |||||  ||| || | ||  ||| |||||||| || ||||||   |
Db   1635 CTGGAAGCTGGCAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACG 1691

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
             || ||    ||||| |||||||| ||||||
Db   1692 TTGTGATACAAGAGATTCCTTTCTTCTTGTGAC 1724
```

```
RESULT  12
HSUDPGT
LOCUS       HSUDPGT      2093 bp    mRNA            PRI       23-MAR-1995
DEFINITION  Human mRNA for liver microsomal UDP-glucuronosyltransferase
            (UDPGT).
ACCESSION   Y00317
VERSION     Y00317.1  GI:37588
KEYWORDS    transferase; UDP-glucuronosyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2093)
  AUTHORS   Burchell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1987) Burchell B., Dept. Biochemistry, Medical
            Sciences Institute, University Dundee, Scotland
REFERENCE   2  (bases 1 to 2093)
  AUTHORS   Jackson,M.R., McCarthy,L.R., Harding,D., Wilson,S., Coughtrie,M.W.
            and Burchell,B.
  TITLE     Cloning of a human liver microsomal UDP-glucuronosyltransferase
            cDNA
  JOURNAL   Biochem. J. 242 (2), 581-588 (1987)
  MEDLINE   87241362
FEATURES             Location/Qualifiers
     source          1. .2093
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /clone_lib="lambda gt11 liver cDNA"
                     /clone="Hlug25"
     sig_peptide     38. .106
                     /product="put. signal peptide"
     CDS             38. .1624
                     /note="precursor"
                     /codon_start=1
                     /protein_id="CAA68415.1"
                     /db_xref="GI:37589"
                     /db_xref="SWISS-PROT:P06133"
                     /translation="MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK
                     DTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELL
                     AELLKIPFVYRPRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVL
                     YFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTRAFITHGGANGIYKAISPRIPMVG
                     VPLFADQPDNIAHMKAKGAAVSLDFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVTGFLLACVATVIFII
                     TKCLFCVWKFVRTGKKGKRD"
     mat_peptide     107. .1621
                     /product="mature UDPGT (AA 1-505) (EC 2.4.1.17)"
     misc_feature    979. .988
                     /note="pot. glycosylation site"
     misc_feature    2076. .2081
                     /note="pot. polyA signal"
BASE COUNT      638 a    397 c    439 g    619 t
ORIGIN
```

Query Match 78.4%; Score 1343; DB 93; Length 2093;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1515; Conservative 0; Mismatches 190; Indels 8; Gaps 5;

```
Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
          || |||||||| ||||||| || |||||||||||||| ||||||||||||||| ||  |
Db     22 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCT-GA 80

Qy     61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
          | |||||||||||||||||||||||||||||| |||||||||||| || |||||| |||||
Db     81 GCTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 140

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
          | |||||||||||| |||||||||||| | || ||||| ||||||||||||||||||||
Db    141 ACTGGATGAATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTG 200

Qy    181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
          || |||||||||||||||||||| ||| ||||||||     |||| ||||||||||||
Db    201 TATTGGCATCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTG 260

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
          ||||||||||  |||||||||||||||||| |||||| |||| |||| ||| | |||||
Db    261 AAGTTTATCCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTA 320

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
          |||||||| ||||  ||| |||||| |  ||||||| |||||||||||||| ||||||||
Db    321 AGAGATGGGCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAA 380

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
          || |||||  |||  |||||||| ||||||| |||||||| ||| ||||||||||||||
Db    381 TCATGTGGACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGA 440

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttttttc 480
          || ||||||| |||||||| ||||||||||||||  | ||| |||||||||||||||| |
Db    441 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCC 500

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
          ||| |||||||||||||| | | ||||||| ||||  ||||| ||||| |  |||||
Db    501 CCTTTGGTGAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGGCCTCGCTTCT 560

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
          |||||||||||  |||||||| ||| ||||||||||  | ||||||||||||||  | |
Db    561 CTCCTGGCTACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGC 620

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
          || |||||||||| || ||||||| |||||||||||||| ||||||||||||||||||||
Db    621 CTGTTGTTATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGA 680

Qy    661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
          |||||||||||||||||| |||||||||||||| | ||| ||||||||||||||||||
```

```
Db    681 TCTATGTGCTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGT 740

Qy    721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
          | ||||||||||| |||||||||||||||| |||| |||||||||| ||||||||||
Db    741 TCTACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACA 800

Qy    781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
          |||||||||| ||||||| |||| |||||||||||||| ||| |||||||||| ||||
Db    801 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTG 860

Qy    841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
          | || ||||||||| |||||| |||||||||||||||||||||| |||||||||| |||
Db    861 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAG 919

Qy    901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
          ||||| ||||||||||||||||||||||||||||||||||||||||||||| || | |||
Db    920 TTTGTCCAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGT 979

Qy    961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
          |||| | ||| |||||||||||||||||||||| ||||||||||||||||||||||||
Db    980 AACACGTCAGAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1039

Qy   1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
          |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||
Db   1040 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTAC 1099

Qy   1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
          |||||||||||||||||||| ||||| ||||| ||||||||||||||||||||||||||
Db   1100 AAGTGGATACCCCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCAT 1159

Qy   1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
          |||||||||||||||||||| ||||||||| | | | ||||||||||||||||| |||||
Db   1160 GGTGGAGCCAATGGCATCTATAAGGCAATCTCTCCTAGAATCCCTATGGTGGGCGTTCCA 1219

Qy   1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
          ||||||   |||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1220 TTGTTTGCAGATCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGT 1279

Qy   1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
          ||||||||| |||||||||||||||||||| | || ||||||||||||||||||||||
Db   1280 TTGGACTTCCACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAAT 1339

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
          |||||||||||||||||||  |||||||||||||||||||| ||||||||||||||
Db   1340 GATCCTTTATATAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTG 1399

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
          |||||||| |||||||||||||||||||||||||||||||||||||| ||| ||||||||||
Db   1400 AAGCCCCTTGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAG 1459

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
          |||||||| |||||||||| |||||||||||||||||||||||||||||||||||||| |
Db   1460 CACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACT 1519

Qy   1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
          ||||| |||||||||||||||||||||||||||||||| ||||||||||    |||||||||
Db   1520 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTT 1576

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
          ||| |||||||||||| ||||| ||| |||||||| |||||||||| ||| ||||||
Db   1577 TGTGTCTGGAAGTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTCTGAGG 1636

Qy   1621 tttgaagctggaaaaccagatagatagacaacttcagtttattccagcaagaaagaaaa 1680
          | |||||||| |||||  ||| || | ||  ||| |||||||| || ||||||   |
Db   1637 CTGGAAGCTGGGAAACCCAATAAAT-GAACTCCTTTAGTTTATTACAACAAGAA--GACG 1693

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
             || ||    ||||| |||||||| ||||||
Db   1694 TTGTGATACAAGAGATTCCTTTCTTCTTGTGAC 1726


RESULT 13
AF112112
```

```
LOCUS       AF112112     2108 bp    mRNA            PRI       11-APR-1999
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase 2B19 precursor,
            mRNA, complete cds.
ACCESSION   AF112112
VERSION     AF112112.1  GI:4580601
KEYWORDS    .
SOURCE      crab-eating macaque.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Molecular cloning, expression and characterization of a monkey
            steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates
            testosterone
  JOURNAL   Eur. J. Biochem. 260 (3), 701-708 (1999)
  MEDLINE   99203465
REFERENCE   2  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2108
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             26. .1612
                     /EC_number="2.4.1.17"
                     /note="UGT2B19"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B19 precursor"
                     /protein_id="AAD24435.1"
                     /db_xref="GI:4580602"
                     /translation="MSMKWTSALLLIQLSCYLSFGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLAYSTSILPDPNNPSPLKFEICPTSLTETEFQDSVTQLVKRWSDIRK
                     DTFWPHFLHVQEMMWTYGDMIRKFCKDVVSNKKLMKKLQESRFDVVLADAISPCGELL
                     AELLKIPFVYSLRFSPGYALEKHGGGFLFPPSYVPVTMSELRDQMTFMERVQNMIYMV
                     YFDFWFQVWDVKNWDQFYSKVLGRPTTLFEIMAKAEIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGDNGVVVFSLGSMVSNMSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTQLYKWLPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVRLDFDTMSSTDLLNALKTVINDPIYKENAMKLSSIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLACVATVIFII
                     TKCLFCVWKFVRTRKKGKRD"
BASE COUNT      649 a    393 c    438 g    628 t
ORIGIN


  Query Match           77.6%;  Score 1329.8;  DB 88;  Length 2108;
  Best Local Similarity  87.6%;  Pred. No. 0;
  Matches 1501;  Conservative    0;  Mismatches  202;  Indels   10;  Gaps    4;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
          || |||||||| |||||| || |||||||||||||| ||||||||||||||| ||  |
Db     10 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCT-GA 68

Qy     61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
          | |||||||| |||| |||||||||||||| |||||||||||| || |||||| |||||
Db     69 GCTGTTACTTGAGCTTTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 128

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
          |||||||||||||| |||||||| ||| | |||||||| ||||||||||||||||||||
Db    129 ATTGGATGAATATAAAGACAATTCTGGATGAGCTTGTCCAGAGAGGTCATGAGGTGACTG 188

Qy    181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
          |||||||||| |||| |||||||||||  |||||||||  |  ||||  ||||||||||||
Db    189 TACTGGCATATTCAACTTCCATTCTTCCTGATCCCAACAACCCATCTCCTCTTAAATTTG 248

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
          || ||| |||||||||||||||  ||||||| ||| || ||| | ||| ||||| |||||
Db    249 AAATTTGTCCTACATCTTTAACCGAAACTGAGTTTCAGGATAGCGTCACGCAACTGGTTA 308

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
```

```
Db    309 AGAGATGGTCAGACATTCGAAAAGACACATTTTGGCCACATTTTTTACACGTACAAGAAA 368

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
Db    369 TGATGTGGACATATGGTGACATGATTAGAAAGTTCTGTAAAGATGTGGTTTCTAATAAGA 428

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
Db    429 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCGGATGCTATTTCCC 488

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
Db    489 CCTGTGGTGAGCTACTGGCTGAGCTACTTAAAATACCCTTTGTGTACAGCCTCCGCTTCT 548

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
Db    549 CTCCTGGCTACGCACTTGAAAAGCATGGTGGAGGGTTTCTGTTCCCTCCTTCCTATGTAC 608

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
Db    609 CTGTTACTATGTCAGAATTAAGAGATCAAATGACTTTCATGGAGAGAGTACAAAATATGA 668

Qy    661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
Db    669 TCTATATGGTTTATTTTGACTTTTGGTTCCAAGTATGGGATGTAAAGAACTGGGATCAGT 728

Qy    721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
Db    729 TTTACAGTAAAGTTCTAGGAAGACCCACTACATTATTTGAGATAATGGCAAAAGCTGAAA 788

Qy    781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
Db    789 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCATCCACTCTTACCAAATGTTG 848

Qy    841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
Db    849 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAG 907

Qy    901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
Db    908 TTCGTCCAGAGCTCTGGAGACAATGGTGTTGTGGTGTTTTCTCTGGGGTCCATGGTCAGT 967

Qy    961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
Db    968 AACATGTCAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1027

Qy   1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
Db   1028 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGACTCAATACTCAGCTGTAC 1087

Qy   1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
Db   1088 AAGTGGCTACCCCAGAATGATCTTCTTGGTCATCCAAAAACCAGAGCTTTTATAACTCAC 1147

Qy   1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
Db   1148 GGTGGAGCCAATGGCATCTACGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCA 1207

Qy   1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
Db   1208 TTGTTTGCCGATCAACCTGATAACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTTAGA 1267

Qy   1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
Db   1268 CTGGACTTCGACACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATAAAT 1327

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
Db   1328 GATCCTATATATAAAGAGAATGCTATGAAATTATCAAGCATTCATCATGATCAACCAGTG 1387

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
Db   1388 AAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAA 1447

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
Db   1448 CACCTTCGGGTTGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATT 1507

Qy   1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
Db   1508 GGGTTCCTGCTGGCCTGTGTGGCAACTGTGATATTCATCATCACAAA---ATGTCTGTTT 1564

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
Db   1565 TGTGTCTGGAAGTTTGTTAGAACAAGAAAGAAGGGAAAAAGAGATTAAATACATCTGAGG 1624

Qy   1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
Db   1625 CTGGAAGCTGGGAAACCTGATAAATGAGATTTCTTCAGTTCATAC-----AATGAGAATA 1679

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
Db   1680 GGTTGTGATACAAGATTCCTTTCTTCTTGTGCC 1712


RESULT  14
AX040087
LOCUS       AX040087     1650 bp    DNA             PAT       18-NOV-2000
DEFINITION  Sequence 10 from Patent WO0063351.
ACCESSION   AX040087
VERSION     AX040087.1  GI:11230049
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Baughn,M.R. and Yang,J.
  TITLE     Carbohydrate-modifying enzymes
  JOURNAL   Patent: WO 0063351-A 10 26-OCT-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2912330CB1"
BASE COUNT      489 a    330 c    354 g    477 t
ORIGIN


  Query Match           71.3%;  Score 1221.4;  DB 9;  Length 1650;
  Best Local Similarity  85.6%;  Pred. No. 9.6e-276;
  Matches 1382;  Conservative    0;  Mismatches  231;  Indels    2;  Gaps    2;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
Db     19 ATTGCATTGCATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTCCTGATACAGCT-GA 77

Qy     61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
Db     78 GCTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCC 137

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
Db    138 ACTGGATGAATATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTG 197

Qy    181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
Db    198 TATTGGCATCTTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTG 257

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
Db    258 AAGTTTATCCTGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTA 317

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
```

```
Db    318 AGAGATGGGCAGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAA 377

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
Db    378 TCATGTGGACATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGA 437

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
Db    438 AACTTATGAAGAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCC 497

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
Db    498 CCTTTGGTGAGCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCT 557

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
Db    558 CTCCTGGCTACGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGC 617

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
Db    618 CTGTTGTTATGTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGA 677

Qy    661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
Db    678 TCTATGTGCTTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGT 737

Qy    721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
Db    738 TCTACAGTGAAGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACA 797

Qy    781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
Db    798 TATGGCTTATTCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTG 857

Qy    841 attttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggag 900
Db    858 AGTTCGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAG 916

Qy    901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
Db    917 TTTGTCCAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGT 976

Qy    961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
Db    977 AACACGTCAGAAGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAG 1036

Qy   1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
Db   1037 GTTCTGTGGAGATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTAC 1096

Qy   1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttataactcat 1140
Db   1097 AAGTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCAT 1156

Qy   1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
Db   1157 GGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCC 1216

Qy   1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
Db   1217 ATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAA 1276

Qy   1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
Db   1277 ATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACC 1336

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
Db   1337 GATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTA 1396

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
Db   1397 AAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAG 1456

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
Db   1457 CACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATT 1516

Qy   1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
Db   1517 GGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTT 1576

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtc 1615
Db   1577 TCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1631
```

```
RESULT  15
HSU59209
LOCUS       HSU59209     2107 bp    mRNA            PRI       02-JUL-1998
DEFINITION  Homo sapiens C19steroid specific UDP-glucuronosyltransferase mRNA,
            complete cds.
ACCESSION   U59209
VERSION     U59209.1  GI:3287472
KEYWORDS    UGT2B17G.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2107)
  AUTHORS   Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
  TITLE     Isolation and characterization of a novel cDNA encoding a human
            UDP-glucuronosyltransferase active on C19 steroids
  JOURNAL   J. Biol. Chem. 271 (37), 22855-22862 (1996)
  MEDLINE   96394358
REFERENCE   2  (bases 1 to 2107)
  AUTHORS   Hum,D.W., Belanger,A., Beaulieu,M. and Levesque,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-1996) Laboratory of Molecular Endocrinology,
            Centre Hospitalier de l'universite Laval, 2705 Boul. Laurier,
            Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2107
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="prostate"
                     /cell_type="LNCaP"
     CDS             52. .1644
                     /note="UGT2B17"
                     /codon_start=1
                     /product="C19steroid specific UDP-glucuronosyltransferase"
                     /protein_id="AAC25491.1"
                     /db_xref="GI:3287473"
                     /translation="MSLKWMSVFLLMQLSCYFSSGSCGKVLVWPTEYSHWINMKTILE
                     ELVQRGHEVIVLTSSASILVNASKSSAIKLEVYPTSLTKNDLEDFFMKMFDRWTYSIS
                     KNTFWSYFSQLQELCWEYSDYNIKLCEDAVLNKKLMRKLQESKFDVLLADAVNPCGEL
                     LAELLNIPFLYSLRFSVGYTVEKNGGGFLFPPSYVPVVMSELSDQMIFMERIKNMIYM
                     LYFDFWFQAYDLKKWDQFYSEVLGRPTTLFETMGKAEMWLIRTYWDFEFPRPFLPNVD
                     FVGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMISNMSEESANMIASALAQIPQ
                     KVLWRFDGKKPNTLGSNTRLYKWLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMV
                     GIPLFADQHDNIAHMKAKGAALSVDIRTMSSRDLLNALKSVINDPIYKENIMKLSRIH
                     HDQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWIQYHSLDVIAFLLACVATMIFM
                     ITKCCLFCFRKLAKTGKKKRD"
BASE COUNT      657 a    382 c    424 g    644 t
ORIGIN
```

```
  Query Match           70.1%;  Score 1201.6;  DB 93;  Length 2107;
  Best Local Similarity  83.6%;  Pred. No. 4.2e-271;
  Matches 1435;  Conservative    0;  Mismatches  269;  Indels   12;  Gaps    6;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
Db     36 ATTGCATAAGACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCT-CA 94
```

```
Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           ||||||||||||||||||||||||||||||| |||||||||||| || ||||||||||||
Db      95 GTTGTTACTTTAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCC 154

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           ||||||| |||||||||||||||| |||||||||||||| |||||||||||| ||
Db     155 ATTGGATAAATATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTG 214

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           |  || ||||||| ||||| |||||| |  || ||| | |  |||| || ||||||| |
Db     215 TGTTGACATCTTCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAG 274

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||||||||||||||||||||||||||| ||| || ||  || |  | ||| ||  | |
Db     275 AAGTTTATCCTACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCG 334

Qy     301 agagatggtcaga---cattcgaaaagatagcttttggttatatttttcacaagaacaag 357
           | |||||| || |    |||  |||| |||  ||||||| |||||||||||||  |||||
Db     335 ATAGATGGACATATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAG 394

Qy     358 aaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaata 417
           || |    |||||||  | ||||     || |||  ||||| |||||  |||||  || |
Db     395 AATTGTGTTGGGAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACA 454

Qy     418 agaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttt 477
           ||||| |||||| |||||||||||||||| ||||||  || || | ||||||||  |||
Db     455 AGAAACTTATGAGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTA 514

Qy     478 ttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgct 537
           |||||||||||||||||||||||  ||||||||||||  ||| ||||||||||||||||
Db     515 ATCCCTGTGGTGAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCT 574

Qy     538 ttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctaca 597
           |  ||  |||||||||| |||| | | |  |||||||| |  | ||||||||||||||
Db     575 TCTCTGTTGGCTACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATG 634

Qy     598 tacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaata 657
           ||||| |||||||||| |||||||||||||||||| |||||||||||| |||||||||
Db     635 TACCTGTTGTTATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATA 694

Qy     658 tgatctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatc 717
           |||| ||| ||||||||||||||||||||| |||   | |||| ||||||||||||| |
Db     695 TGATATATATGCTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACC 754

Qy     718 agttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctg 777
           ||||||| |||||||| |||||||||||||| ||||||||||||||| |||||||
Db     755 AGTTTTATAGTGAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTG 814

Qy     778 acatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacg 837
           | || ||||| || |||| ||  |||  |||| |||||||||  |||||||||||||| |
Db     815 AAATGTGGCTCATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATG 874

Qy     838 ttgattttgttggaggattccactggcaaacctgccaaaacccctacctaaggaaatggag 897
           |||||||||||||||| | |||| | ||||| |||||||| | ||||||||||||||
Db     875 TTGATTTTGTTGGAGGACTTCACT-GTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAA 933

Qy     898 gagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgata 957
           |||||||| ||||||||||||||||||||| ||||||||||||||||||||  ||||
Db     934 GAGTTTGTGCAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATC 993

Qy     958 agtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaa 1017
           ||||||||| ||| |||||| |||||  | |||||| |||||||||| ||||||||||
Db     994 AGTAACATGTCAGAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAA 1053

Qy    1018 aaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctg 1077
           |||||||| |||||||||||||| || || ||| || | ||||||  |||||||| |||
Db    1054 AAGGTTCTATGGAGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTG 1113

Qy    1078 tataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataact 1137
           ||||||||| |||||||||||||||||||| |||||||| ||||||| ||||||||||||
Db    1114 TATAAGTGGTTACCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACT 1173

Qy    1138 catggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcatt 1197
           |||||||| |||||||||||||||||||| |||||||||||| |||||||||||||||||
Db    1174 CATGGTGGAACCAATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATT 1233

Qy    1198 ccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgtt 1257
           || ||||||  |||||||| |||||||||||||||||||||| ||||||||||||| |
Db    1234 CCCTTGTTTGCGGATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTC 1293

Qy    1258 agattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaatt 1317
           ||  ||||| ||| || ||||| |||| ||| |||| |||||| ||||| |||| |||
Db    1294 AGTGTGGACATCAGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATT 1353

Qy    1318 aatgatccttttatataaagagaatattatgaaattatcaagaattcaacatgatcaacca 1377
           ||||| ||| | |||||||||||||| |||||||||||||||||| ||||||||||
Db    1354 AATGACCCTATCTATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCG 1413

Qy    1378 gtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagcc 1437
           || ||||||||||||||||||||||||||||||||||| |||||||||| ||| |||||||||
Db    1414 GTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCC 1473

Qy    1438 aaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtg 1497
           || |||||||| || |||||||| |||||||||| ||||||||||||||||||||||||
Db    1474 AAGCACCTTCGGGTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTG 1533

Qy    1498 attgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctg 1557
           || |  || |||||||||| ||||||||| |||||||||  |||||||| | ||| |||
Db    1534 ATAGCATTCCTGCTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTG 1593

Qy    1558 ttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1617
           ||||||||| | ||| |||| | || ||| ||||||   |||| ||||||||| ||
Db    1594 TTTTGTTTCCGAAAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAA 1653

Qy    1618 acatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaaga 1677
           |    |||| |||||   |  | |||| ||||  || |  |||||||||||| | | |
Db    1654 AAGCCTGAAG-TGGAATGACCAAAAGATGGGACTCCTCC--TTTATTCCAGCATGGAGG- 1709

Qy    1678 aaagattgttatgcaagatttctttcttcctgtgac 1713
           | ||   ||| | |||||| || |||||| |||
Db    1710 ---GTTTTAAATGGAGGATTTCCTTTTTCCTGCGAC 1742
```

Search completed: April 30, 2001, 18:55:22
Job time: 36196 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:04:18 ; Search time 933.78 Seconds
(without alignments)
1070.935 Million cell updates/sec

Title: US-09-721-183-2
Perfect score: 1713
Sequence: 1 atcgcattgcaccaggatga.....gatttctttcttcctgtgac 1713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1443.6	84.3	1854	21	Z95200	Human UDP-glucuron
2	1354.2	79.1	2092	21	Z95199	Human UDP-glucuron
3	1221.4	71.3	1650	21	C65396	Human carbohydrate
4	1201.6	70.1	2107	19	V15900	Uridine diphospho-
5	1188.8	69.4	1976	21	Z95206	Human UDP-glucuron
6	602.6	35.2	1686	21	Z95201	Human UGT2B7 exon
7	579.2	33.8	1323	21	Z95193	Human UGT2B4 exon
8	491.4	28.7	2312	21	Z95207	Human UGT2B15 exon
9	326.6	19.1	596	21	Z95205	Human UGT2B7 exon
10	311.4	18.2	2368	13	Q27370	HUG-Br2. Homo sap
11	294.8	17.2	2351	13	Q27369	HUG-Br1. Homo sap
12	266.2	15.5	1589	21	Z95198	Human UGT2B4 exon
13	265.8	15.5	735	21	Z45118	UDP-glucuronosyltr
14	246	14.4	978	21	Z95211	Human UGT2B15 exon
15	220.6	12.9	350	21	C03286	Human secreted pro
16	207.6	12.1	1591	21	Z95204	Human UGT2B7 exon
17	205.8	12.0	689	21	Z95197	Human UGT2B4 exon
18	181.6	10.6	1602	21	Z95210	Human UGT2B15 exon
19	166	9.7	264	21	F15707	Human prostate can
20	156	9.1	223	21	A87255	Rat hepatocyte car
21	126.2	7.4	391	20	Z07208	Human lung tumour
22	126.2	7.4	391	21	C79066	Human lung tumour-
23	124.4	7.3	2320	21	C58107	Human PRO1780 nucl
24	124.4	7.3	2320	21	A37101	Human PRO1780 (UNQ
25	124	7.2	785	21	Z95195	Human UGT2B4 exon
26	122.4	7.1	1340	21	Z95202	Human UGT2B7 exon
27	121	7.1	1822	21	Z95203	Human UGT2B7 exon
28	120.6	7.0	746	21	Z95194	Human UGT2B4 exon
29	120	7.0	273	21	A87491	Rat hepatocyte car
30	114.6	6.7	480	21	Z95209	Human UGT2B15 exon
31	111	6.5	1020	21	Z95208	Human UGT2B15 exon
32	108	6.3	283	21	A87467	Rat hepatocyte car
33	106.6	6.2	1667	13	Q33027	Exon 5 from the UG
34	100.6	5.9	515	20	V87412	EST clone BR77. H
35	100.2	5.8	164	21	A87154	Rat hepatocyte car
36	96.4	5.6	508	21	C98478	Human colon cancer
37	91.6	5.3	259	21	A87473	Rat hepatocyte car
38	84.2	4.9	1138	21	Z95196	Human UGT2B4 exon
39	81.6	4.8	775	13	Q33026	Exons 3-4 from the
40	74.8	4.4	219	21	C13974	Human secreted pro
41	71.6	4.2	366	21	A87085	Rat hepatocyte car
42	60.8	3.5	1048	13	Q33020	UGT1E Exon 1 from
43	60.4	3.5	867	21	Z45111	UDP-glucuronosyltr
44	60.4	3.5	867	21	Z45112	UDP-glucuronosyltr
45	59.2	3.5	867	21	Z45113	UDP-glucuronosyltr

ALIGNMENTS

RESULT 1
Z95200
ID Z95200 standard; DNA; 1854 BP.
XX
AC Z95200;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B7 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; Y78934.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX

PS Disclosure; Page 41-44; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B7
CC (UGT2B7) gene. UDP-glucuronosyltransferase (UGTs) are a family of enzymes
CC that catalyse the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC substrate in an individual. The nucleic acid molecules comprising a human
CC UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of UGT2B
CC substrate, potential drug-drug interactions and adverse side effects. The
CC polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC detecting genetic linkage related to phenotypic variation in activity or
CC expression of UGT2B protein. The polymorphism containing nucleic acid
CC molecules may also be used for generating genetically modified non-human
CC animals and for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 1854 BP; 572 A; 338 C; 392 G; 552 T; 0 other;

Query Match 84.3%; Score 1443.6; DB 21; Length 1854;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 1564; Conservative 0; Mismatches 139; Indels 7; Gaps 3;

```
Qy      4 gcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccagtt 63
          ||||||||||||||||| || |||||||||||||||||  || |||| ||||| ||  || |
Db      2 gcattgcaccaggatgtctgtgaaatggacttcagtaattttgctaatacaactg-agct 60

Qy     64 gttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccatt 123
            || ||||||||||||| |||||||| |||||||||||||| ||||||||||||||||||
Db     61 tttgctttagctctgggaattgtggaaaggtgctggtgtgggcagcagaatacagccatt 120

Qy    124 ggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtac 183
          |||||||||| |||||||||| | |||||| ||||||||||||||||||||||||||||
Db    121 ggatgaatataaagacaatcctggatgagcttattcagagaggtcatgaggtgactgtac 180

Qy    184 tggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaag 243
          |||||||||||||||||||||||||||||||  |  ||||| |||||||| |||||
Db    181 tggcatcttcagcttccattctttttgatcccaacaactcatccgctcttaaaattgaaa 240

Qy    244 tttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaaga 303
          ||||||| |||||||||||||||||||| || |||||| |||||||||||||| ||||||
Db    241 tttatcccacatctttaactaaaactgagttggagaatttcatcatgcaacagattaaga 300

Qy    304 gatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcc 363
          |||||||||||  ||| ||||||||  |||||||||||||||||||||||  ||| ||||||
Db    301 gatggtcagaccttccaaaagatacattttggttatatttttcacaagtacaggaaatca 360

Qy    364 tgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaag 423
          |||    |||   |||||||  |||||| ||||||||||||||||||||||||||||||
Db    361 tgtcaatatttggtgacataactagaaagttctgtaaagatgtagtttcaaataagaaat 420

Qy    424 ttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccct 483
          |||||||||| |||||||||||||||||||| || |||||||||||||||| |||||||||
Db    421 ttatgaaaaaagtacaagagtcaagatttgacgtcatttttgcagatgctatttttccct 480

Qy    484 gtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactc 543
          || |||||||||||||| |||| |||||||||  ||||||||||||||| ||||  |||
Db    481 gtagtgagctgctggctgagctatttaacataccctttgtgtacagtctcagcttctctc 540

Qy    544 ctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacataccta 603
          ||||||||||  |||||| ||| ||||||||| | ||||||||||||||||| ||||||
Db    541 ctggctacacttttgaaaagcatagtggaggatttatttccctccttcctacgtacctg 600

Qy    604 ttgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatct 663
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||
Db    601 ttgttatgtcagaattaactgatcaaatgactttcatggagagggtaaaaaatatgatct 660

Qy    664 atgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagtttt 723
          ||||||||| |||||||||||||| |||| | ||| ||||||||||||||||||||||
Db    661 atgtgctttactttgacttttggttcgaaatatttgacatgaagaagtgggatcagtttt 720

Qy    724 acagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatat 783
          | ||||||||| |||||||||||||||| |||| |||||||||| |||||||| |||
Db    721 atagtgaagttctaggaagacccactacattatctgagacaatggggaaagctgacgtat 780

Qy    784 ggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgatt 843
          ||||||| |||||||||||||| |||||| ||||| |||||  |||||||||| |||||||
Db    781 ggcttattcgaaactcctggaattttcagtttccatatccactcttaccaaatgttgatt 840

Qy    844 ttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggagttt 903
          |||||||||||  |||||| |||||||||||||||||||  |||||||||||||| || |||
Db    841 ttgttggaggactccact-gcaaacctgccaaaccccctgcctaaggaaatggaagacttt 899

Qy    904 gtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaac 963
          ||||||||||||||||||||||||||||||||||||||||||||||||||| || | ||||||
Db    900 gtacagagctctggagaaaatggtgttgtggtgttttctctggggtcaatggtcagtaac 959

Qy    964 atgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggtt 1023
          ||||||| |||||||||||| |||||||| ||||||| ||| ||||||||||||||||||
Db    960 atgacagaagaaagggccaacgtaattgcatcagccctggcccagatcccacaaaaggtt 1019

Qy   1024 ctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataag 1083
          |||||||||||||||||||||||||||||||||||||||||||||||||| || |||||||
Db   1020 ctgtggagatttgatgggaataaaccagataccttaggtctcaatactcggctctacaag 1079

Qy   1084 tggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggt 1143
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1080 tggataccccagaatgaccttctaggtcatccaaagaccagagcttttataactcatggt 1139

Qy   1144 ggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattg 1203
          |||||||||||||||| |||||||||||||||||||||||||||||||| |||||||||
Db   1140 ggagccaatggcatctacgaggcaatctaccatgggatccctatggtggggattccattg 1199

Qy   1204 ttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattg 1263
          |||   ||||||||||||||||||||||||||||||||||||||| |||||||||||| ||
Db   1200 tttgccgatcaacctgataacattgctcacatgaaggccaggggagcagctgttagagtg 1259

Qy   1264 gacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgat 1323
          |||||||||||||||||||||||||||| |||||||||| |||||| ||||||||||||
Db   1260 gacttcaacacaatgtcgagtacagacttgctgaatgcattgaagagagtaattaatgat 1319

Qy   1324 cctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaag 1383
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||||| |||
Db   1320 ccttcatataaagagaatgttatgaaattatcaagaattcaacatgatcaaccagtgaag 1379

Qy   1384 cccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacac 1443
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||
Db   1380 cccctggatcgagcagtcttctggattgaatttgtcatgcgccacaaaggagctaaacac 1439

Qy   1444 cttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattggg 1503
          ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1440 cttcgggttgcagcccacgacctcacctggttccagtaccactctttggatgtgattggg 1499

Qy   1504 tttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgt 1563
          || ||||||| |||||||||||||||||||||||||||||||| ||||||| | ||||||||||
Db   1500 ttcctgctggtctgtgtggcaactgtgatatttatcgtcacaaaatgttgtctgttttgt 1559

Qy   1564 ttctggaagtttgctagaaaaagggaagaagggaaaaagagattagttatgtctgacattt 1623
          |||||||||||||||||||||  |||||||||||||  |||||||||| ||||| ||||
Db   1560 ttctggaagtttgctagaaaagcaaagaagggaaaaaatgattagttatatctgagattt 1619

Qy   1624 gaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagat 1683
          |||||||||||||| ||||| |  ||| ||||||||||||||||||||  .  ||||||
Db   1620 gaagctggaaaacctgataggtgagactacttcagtttattccagcaag-----aaagat 1674

Qy   1684 tgttatgcaagatttctttcttcctgtgac 1713
          ||| |||||||||||||||||||| |||
Db   1675 tgtgatgcaagatttctttcttcctgagac 1704
```

RESULT 2

```
Z95199
ID   Z95199 standard; DNA; 2092 BP.
XX
AC   Z95199;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B4 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   P-PSDB; Y78933.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions -
XX
PS   Disclosure; Page 34-36; 72pp; English.
XX
CC   This sequence represents the human UDP-glucuronosyltransferase 2B4
CC   (UGT2B4) gene. UDP-glucuronosyltransferase (UGTs) are a family of enzymes
CC   that catalyse the glucuronic acid conjugation of a wide range of
CC   endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC   steroid metabolizing isoforms in the liver. Alteration of the expression
CC   or function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC   locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC   substrate in an individual. The nucleic acid molecules comprising a human
CC   UGT2B sequence polymorphism can be used in screening assays for
CC   genotyping individuals, also to predict their rate of metabolism of UGT2B
CC   substrate, potential drug-drug interactions and adverse side effects. The
CC   polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC   detecting genetic linkage related to phenotypic variation in activity or
CC   expression of UGT2B protein. The polymorphism containing nucleic acid
CC   molecules may also be used for generating genetically modified non-human
CC   animals and for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 2092 BP; 639 A; 398 C; 438 G; 617 T; 0 other;

  Query Match            79.1%;  Score 1354.2;  DB 21;  Length 2092;
  Best Local Similarity  88.8%;  Pred. No. 0;
  Matches 1522;  Conservative    0;  Mismatches  183;  Indels    8;  Gaps    5;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
Db      22 attgcattgcatcaggatgtctatgaaatggacttcagctcttctgctgatacagct-ga 80

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
Db      81 gctgttactttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagcc 140

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
Db     141 actggatgaatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactg 200

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
Db     201 tattggcatcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttg 260

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
Db     261 aagtttatcctgtatctttaactaaaactgagtttgaggatattatcaagcagctggtta 320

Qy     301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
Db     321 agagatgggcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaa 380

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
Db     381 tcatgtggacatttaatgacatacttagaaagttctgtaaggatatagtttcaaataaga 440

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
Db     441 aacttatgaagaaactacaggagtcaagatttgatgttgttcttgcagatgctgttttcc 500

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
Db     501 cctttggtgagctgctggccgagttacttaaaataccctttgtctacagcctccgcttct 560

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
Db     561 ctcctggctacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgc 620

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
Db     621 ctgttgttatgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatga 680

Qy     661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
Db     681 tctatgtgctttattttgaattttggttccaaatatttgacatgaagaagtgggatcagt 740

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
Db     741 tctacagtgaagttctaggaagacccactacgttatctgagacaatggcaaaagctgaca 800

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
Db     801 tatggcttattcgaaactactgggattttcaatttcctcacccactcttaccaaatgttg 860

Qy     841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
Db     861 agttcgttggaggactccact-gcaaacctgccaaaccccctaccgaaggaaatggaagag 919

Qy     901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
Db     920 tttgtccagagctctggagaaaatggtgttgtggtgttttctctggggtcgatggtcagt 979

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
Db     980 aacacgtcagaagaaagggccaatgtaattgcatcagcccttgccaagatcccacaaaag 1039

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
Db    1040 gttctgtggagatttgatgggaataaaccagatactttaggactcaatactcggctgtac 1099

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttataactcat 1140
Db    1100 aagtggataccccagaatgatcttcttggtcacccaaaaaccagagctttataactcat 1159

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
Db    1160 ggtggagccaatggcatctatgaggcaatctaccatggaatccctatggtgggcgttcca 1219

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
Db    1220 ttgtttgcagatcaacctgataacattgcacacatgaaggccaagggagcagctgttagt 1279

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
```

```
Db   1280 ttggacttccacacaatgtcgagtacagacttactcaatgcactgaagacagtaattaat 1339

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
Db   1340 gatcctttatataaagagaatgctatgaaattatcaagaattcatcatgatcaaccagtg 1399

Qy   1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
Db   1400 aagccccttgatcgagcagtcttctggattgaatttgtcatgcgccataaaggagccaag 1459

Qy   1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
Db   1460 caccttcgggttgcagcccacgacctcacctggttccagtaccactctttggatgtgact 1519

Qy   1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
Db   1520 gggttcctgctggcctgtgtggcaactgtgatattcatcatcacaaa---atgtctgttt 1576

Qy   1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgaca 1620
Db   1577 tgtgtctggaagtttgttagaacaggaaagaagggaaaagagattaattacgtctgagg 1636

Qy   1621 tttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaa 1680
Db   1637 ctggaagctgggaaacccaataaat-gaactcctttagtttattacaacaagaa--gacg 1693

Qy   1681 gattgttatgcaagatttctttcttcctgtgac 1713
Db   1694 ttgtgatacaagagattcctttcttcttgtgac 1726
```

```
RESULT   3
C65396
ID   C65396 standard; cDNA; 1650 BP.
XX
AC   C65396;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human carbohydrate-modifying enzyme cDNA Incyte ID No: 2912330CB1.
XX
KW   Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW   immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW   antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW   nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW   antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW   dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW   fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW   autoimmune disorder; inflammatory disorder; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200063351-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-US10882.
XX
PR   21-APR-1999;   99US-0130383.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX
DR   WPI; 2000-672729/65.
DR   P-PSDB; B28677.
XX
PT   Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT   for diagnosis, treatment, and prevention of carbohydrate metabolism
PT   disorders, autoimmune/inflammatory disorders, and cancer
XX
PS   Claim 4; Page 75; 75pp; English.
XX
CC   The present cDNA sequence encodes a human carbohydrate-modifying enzyme
CC   (CME). CME polynucleotides and polypeptides are useful for treating and
CC   diagnosing diseases associated with CME such as diabetes,
CC   autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC   adult respiratory distress syndrome, allergies, anaemia, asthma,
CC   atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC   contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC   glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC   Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC   lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC   complications of cancer, haemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC   trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC   useful for drug screening.
XX
SQ   Sequence 1650 BP; 489 A; 330 C; 354 G; 477 T; 0 other;
```

```
  Query Match             71.3%;  Score 1221.4;  DB 21;  Length 1650;
  Best Local Similarity   85.6%;  Pred. No. 0;
  Matches 1382;  Conservative    0;  Mismatches  231;  Indels    2;  Gaps    2;

Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
Db     19 attgcattgcatcaggatgtctatgaaatggacttcagctcttctcctgatacagct-ga 77

Qy     61 gttgttacttttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
Db     78 gctgttacttttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagcc 137

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
Db    138 actggatgaatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactg 197

Qy    181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
Db    198 tattggcatcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttg 257

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
Db    258 aagtttatcctgtatctttaactaaaactgagtttgaggatattatcaagcagctggtta 317

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
Db    318 agagatgggcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaa 377

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
Db    378 tcatgtggacatttaatgacatacttagaaagttctgtaaggatatagtttcaaataaga 437

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
Db    438 aacttatgaagaaactacaggagtcaagatttgatgttgttcttgcagatgctgttttcc 497

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
Db    498 cctttggtgagctgctggccgagttacttaaaataccctttgtctacagcctccgcttct 557

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
Db    558 ctcctggctacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgc 617

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
Db    618 ctgttgttatgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatga 677

Qy    661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
Db    678 tctatgtgctttattttgaattttggttccaaatatttgacatgaagaagtgggatcagt 737
```

```
Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgaca 780
           | ||||||||||| |||||||||||||||| |||| |||||||||| ||||||||||
Db     738 tctacagtgaagttctaggaagacccactacgttatctgagacaatggcaaaagctgaca 797

Qy     781 tatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttg 840
           |||||||||| ||||||| |||| |||||||||||||| ||| |||||||||| ||||
Db     798 tatggcttattcgaaactactgggattttcaatttcctcacccactcttaccaaatgttg 857

Qy     841 attttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggag 900
           | || ||||||||| |||||| |||||||||||||||||||| |||||||||| |||
Db     858 agttcgttggaggactccact-gcaaacctgccaaaccccctaccgaaggaaatggaagag 916

Qy     901 tttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagt 960
           ||||| ||||||||||||||||||||||||||||||||||||||||||  || | |||
Db     917 tttgtccagagctctggagaaaatggtgttgtggtgttttctctggggtcgatggtcagt 976

Qy     961 aacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaag 1020
           |||| | ||| |||||||||||||||||||| ||||||||||||||||||||||||||
Db     977 aacacgtcagaagaaagggccaatgtaattgcatcagcccttgccaagatcccacaaaag 1036

Qy    1021 gttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtat 1080
           |||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||
Db    1037 gttctgtggagatttgatgggaataaaccagatactttaggactcaatactcggctgtac 1096

Qy    1081 aagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcat 1140
           |||||||||||||||||||| ||||| |||||||| ||||||| |||||||| ||||||
Db    1097 aagtggataccccagaatgatcttcttggtcatcccaaaaccaaagcttttatcactcat 1156

Qy    1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           ||||||   ||||| |||||||| || || ||||||||| ||||||||||||  ||||
Db    1157 ggtggaatgaatgggatctatgaagctatttaccatggggtccctatggtgggagttccc 1216

Qy    1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
            | ||| | ||||| | ||||||||| |||||||||||||||| ||||||||||   |
Db    1217 atatttggtgatcagcttgataacatagctcacatgaaggccaaaggagcagctgtagaa 1276

Qy    1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
            |  ||||||| || ||| | ||   ||| | |||| || ||| ||||| ||||
Db    1277 ataaacttcaaaactatgacaagcgaagatttactgagggctttgagaacagtcattacc 1336

Qy    1321 gatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagta 1380
           ||| | |  |||||||||||  ||||| |||||||||||||| ||||||||||| |||
Db    1337 gattcctcttataaagagaatgctatgagattatcaagaattcaccatgatcaacctgta 1396

Qy    1381 aagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaa 1440
           |||||||| |||||||||||||||||||| || |||||||||| ||||||||||||||||
Db    1397 aagccccctagatcgagcagtcttctggatcgagtttgtcatgcgccacaaaggagccaag 1456

Qy    1441 caccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgatt 1500
           ||||| |||   || |||||||||||||||||||||||||| || ||||| | |||||||||
Db    1457 cacctgcgatcagctgcccatgacctcacctggttccagcactactctatagatgtgatt 1516

Qy    1501 gggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttt 1560
           ||||| |||||| ||||||||||||||  |||||  |  |||||||  | ||  | | |||
Db    1517 gggttcctgctgacctgtgtggcaactgctatattcttgttcacaaaatgtttttattt 1576

Qy    1561 tgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtc 1615
           |  |     || |||  || ||   | ||||  | ||| || || ||| | | ||
Db    1577 tcctgtcaaaaatttaataaaactagaaagatagaaaagagggaatagatctttc 1631


RESULT   4
V15900
ID   V15900 standard; cDNA; 2107 BP.
XX
AC   V15900;
XX
DT   26-MAY-1998  (first entry)
XX
DE   Uridine diphospho-glucuronosyltransferase 2B17 (UGT2B17) encoding cDNA.
XX
KW   Uridine diphospho-glucuronosyltransferase 2B17; UGT2B17; catalyse;
KW   androsterone; androsterone-glucuronic acid; androgen; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..51
FT                   /*tag= a
FT   CDS             52..1644
FT                   /*tag= b
FT                   /product= "UGTB17 enzyme"
FT   3'UTR           1645..2107
FT                   /*tag= c
XX
PN   WO9744466-A1.
XX
PD   27-NOV-1997.
XX
PF   16-MAY-1997;   97WO-CA00328.
XX
PR   17-MAY-1996;   96US-0649319.
XX
PA   (ENDO-) ENDORECHERCHE INC.
XX
PI   Beaulieu M,  Belanger A,  Hum DW,  Levesque E;
XX
DR   WPI; 1998-018520/02.
DR   P-PSDB; W47126.
XX
PT   DNA encoding uridine di:phospho:glucuronosyl:transferase 2B17 -
PT   which catalyses conversion of androsterone to
PT   androsterone-glucuronic acid
XX
PS   Claim 15; Pages 4-6; 53pp; English.
XX
CC   This cDNA encodes an enzyme uridine di-phosphoglucuronosyltransferase
CC   2B17 (UGT2B17). This novel enzyme catalyses the conversion of
CC   androsterone to androsterone-glucuronic acid. The UGT2B17 can be used to
CC   detect anti-UGT2B17 antibodies. The antibody can be used to detect a
CC   localised concentration of UGT2B17 or an alteration in androgen activity.
CC   The UGT2B17 can also be used to alter the concentration of an androgenic
CC   compound in a tissue, specifically dihydrotestosterone. An isolated
CC   nucleotide sequence comprising at least 30 consecutive nucleotides from
CC   the coding region of the 2107 base pair sequence, or its complement can
CC   be used to block the synthesis of UGT2B17, e.g. an expression disrupting
CC   sense or antisense fragment, or as a probe for a UGT2B17 coding sequence.
XX
SQ   Sequence 2107 BP; 657 A; 382 C; 424 G; 644 T; 0 other;


  Query Match             70.1%;  Score 1201.6;  DB 19;  Length 2107;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1435;  Conservative    0;  Mismatches  269;  Indels   12;  Gaps    6;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
             || ||||   ||||||||| |||||||||||  |||||  |||||||||| || || ||
Db      36 attgcataagaccaggatgtctctgaaatggatgtcagtctttctgctgatgcagct-ca 94

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           ||||||||||||||||||||||||||||||| |||||||||||| || ||||||||||||
Db      95 gttgttactttagctctgggagttgtggaaaggtgctggtgtggcccacagaatacagcc 154

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           ||||||| |||||||||||||||||| |||||||||||||||| ||||||||||||| ||
Db     155 attggataaatatgaagacaatcctggaagagcttgttcagaggggtcatgaggtgattg 214

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           |  || ||||||| ||||| |||||| |  || ||| | |  ||||  || ||||||| |
Db     215 tgttgacatcttcggcttctattcttgtcaatgccagtaaatcatctgctattaaattag 274

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           |||||||||||||||||||||||||| ||| || ||  || |  | ||| ||  | |
Db     275 aagtttatcctacatctttaactaaaaatgatttggaagatttttttatgaaaatgttcg 334

Qy     301 agagatggtcaga---cattcgaaaagatagcttttggttatatttttcacaagaacaag 357
```

```
Db    335 atagatggacatatagtatttcaaaaaatacattttggtcatatttttcacaactacaag 394

Qy    358 aaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaata 417
Db    395 aattgtgttgggaatattctgactataatataaagctctgtgaagatgcagttttgaaca 454

Qy    418 agaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttt 477
Db    455 agaaacttatgagaaaactacaagagtcaaaatttgatgtccttctggcagatgccgtta 514

Qy    478 ttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgct 537
Db    515 atccctgtggtgagctgctggctgaactacttaacatacccttctgtacagtctccgct 574

Qy    538 ttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctaca 597
Db    575 tctctgttggctacacagttgagaagaatggtggaggatttctgttccctccttcctatg 634

Qy    598 tacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaata 657
Db    635 tacctgttgttatgtcagaattaagtgatcaaatgattttcatggagaggataaaaaata 694

Qy    658 tgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatc 717
Db    695 tgatatatatgctttattttgacttttggtttcaagcatatgatctgaagaagtgggacc 754

Qy    718 agttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctg 777
Db    755 agttttatagtgaagttctaggaagacccactacattatttgagacaatggggaaagctg 814

Qy    778 acatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacg 837
Db    815 aaatgtggctcattcgaacctattgggattttgaatttcctcgcccattcttaccaaatg 874

Qy    838 ttgattttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggag 897
Db    875 ttgattttgttggaggacttcact-gtaaaccagccaaaccccttgcctaaggaaatggaa 933

Qy    898 gagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgata 957
Db    934 gagtttgtgcagagctctggagaaaatggtattgtggtgttttctctggggtcgatgatc 993

Qy    958 agtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaa 1017
Db    994 agtaacatgtcagaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaa 1053

Qy   1018 aaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctg 1077
Db   1054 aaggttctatggagatttgatggcaagaagccaaatactttaggttccaatactcgactg 1113

Qy   1078 tataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataact 1137
Db   1114 tataagtggttaccccagaatgaccttcttggtcatcccaaaaccaaagcttttataact 1173

Qy   1138 catggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcatt 1197
Db   1174 catggtggaaccaatggcatctatgaggcgatctaccatgggatccctatggtgggcatt 1233

Qy   1198 ccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgtt 1257
Db   1234 cccttgtttgcggatcaacatgataacattgctcacatgaaagccaagggagcagccctc 1293

Qy   1258 agattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaatt 1317
Db   1294 agtgtggacatcaggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcatt 1353

Qy   1318 aatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaacca 1377
Db   1354 aatgaccctatctataaagagaatatcatgaaattatcaagaattcatcatgatcaaccg 1413

Qy   1378 gtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagcc 1437
Db   1414 gtgaagcccctggatcgagcagtcttctggattgagtttgtcatgcgccataaaggagcc 1473

Qy   1438 aaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtg 1497
Db   1474 aagcaccttcgggtcgcagcccacaacctcacctggatccagtaccactctttggatgtg 1533

Qy   1498 attgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctg 1557
Db   1534 atagcattcctgctggcctgcgtggcaactatgatatttatgatcacaaaatgttgcctg 1593

Qy   1558 ttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctg 1617
Db   1594 ttttgtttccgaaagcttgccaaaacaggaaagaagaagaaaagggattagttatatcaa 1653

Qy   1618 acatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaagaaaga 1677
Db   1654 aagcctgaag-tggaatgaccaaaagatgggactcctcc--tttattccagcatggagg- 1709

Qy   1678 aaagattgttatgcaagatttctttcttcctgtgac 1713
Db   1710 ---gttttaaatggaggatttccttttttcctgcgac 1742
```

RESULT 5
Z95206
ID Z95206 standard; DNA; 1976 BP.
XX
AC Z95206;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B15 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW drug interaction; detect; human; single nucleotide polymorphism;
KW SNPs; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; Y78935.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 56-59; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B15
CC (UGT2B15) gene. UDP-glucuronosyltransferase (UGTs) are a family of
CC enzymes that catalyse the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see Z95051-Z95110). Probes which detect the UGT2B
CC locus polymorphisms can be used to detect altered UGT2B metabolism of a
CC substrate in an individual. The nucleic acid molecules comprising a human
CC UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of UGT2B
CC substrate, potential drug-drug interactions and adverse side effects. The

```
CC   polymorphisms can be used as single nucleotide polymorphisms (SNPs) for
CC   detecting genetic linkage related to phenotypic variation in activity or
CC   expression of UGT2B protein. The polymorphism containing nucleic acid
CC   molecules may also be used for generating genetically modified non-human
CC   animals and for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 1976 BP; 594 A; 368 C; 419 G; 595 T; 0 other;


  Query Match             69.4%;  Score 1188.8;  DB 21;  Length 1976;
  Best Local Similarity   83.1%;  Pred. No. 0;
  Matches 1417;  Conservative    0;  Mismatches  277;  Indels   12;  Gaps    5;

Qy      11 accaggatgactctgaaatggacttcagttcttctgctgatacatctccagttgttactt 70
           ||||||||| |||||||||||||| |||||  |||||||||||||| || ||||||||||||
Db       5 accaggatgtctctgaaatggacgtcagtctttctgctgatacagct-cagttgttactt 63

Qy      71 tagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaa 130
           |||||||| || |||||||| ||||| |||||| || ||||||||||||||||||| ||
Db      64 tagctctggaagctgtggaaaggtgctagtgtggcccacagaatacagccattggataaa 123

Qy     131 tatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatc 190
           |||||||||||||||| |||||||||||||||| |||||||||||||||  || ||||
Db     124 tatgaagacaatcctggaagagcttgttcagaggggtcatgaggtgactgtgttgacatc 183

Qy     191 ttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatcc 250
           ||| ||||| | |||| |  || ||| | |  |||| || ||||||| ||||||||||
Db     184 ttcggcttctactcttgtcaatgccagtaaatcatctgctattaaattagaagtttatcc 243

Qy     251 tacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg-- 308
           |||||||||||||||| ||| ||.||  ||    |  || ||    |  | ||||||
Db     244 tacatctttaactaaaaatgatttggaagattctcttctgaaaattctcgatagatggat 303

Qy     309 -tcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtg 367
            |  |   ||  |||| |||  ||||||| ||||||||||||  |||||||  |    ||
Db     304 atatggtgtttcaaaaaatacattttggtcatatttttcacaattacaagaattgtgttg 363

Qy     368 ggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttat 427
           |||||  ||||||     ||  ||  |||||||||||| |||||  ||||||||| ||||
Db     364 ggaatattatgactacagtaacaagctctgtaaagatgcagttttgaataagaaacttat 423

Qy     428 gaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtgg 487
           ||  |||||||||||||||  |||||  || || | ||||||||  ||  |||||||||
Db     424 gatgaaactacaagagtcaaagtttgatgtcattctggcagatgcccttaatccctgtgg 483

Qy     488 tgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctgg 547
           ||||||  |||||||  |||  |||||||||  ||| ||||||||||  ||  ||  |||
Db     484 tgagctactggctgaactatttaacatacccttttctgtacagtcttcgattctctgttgg 543

Qy     548 ctacacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgt 607
           |||||||  |||| | | |  |||||||| |  | ||||||||||||||  |||||  ||||
Db     544 ctacacatttgagaagaatggtggaggatttctgttccctccttcctatgtacctgttgt 603

Qy     608 tatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaaatatgatctatgt 667
           |||||||  |||||||||||||||||  |||||||||||||  |||||||||||||  || |
Db     604 tatgtcagaattaagtgatcaaatgattttcatggagaggataaaaaaatatgatacatat 663

Qy     668 gctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacag 727
           ||||||||||||||||||| ||||| | |||| ||||||||||||  ||||||||  ||
Db     664 gctttattttgacttttggtttcaaatttatgatctgaagaagtgggaccagttttatag 723

Qy     728 tgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggct 787
           |||||||  |||||||||||||||||  |||||||||||||||||  ||||||||  || |||||
Db     724 tgaagttctaggaagacccactacattatttgagacaatggggaaagctgaaatgtggct 783

Qy     788 tatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgt 847
            ||  |||| ||  |||  |||| |||||||||  |||||||||||||| ||||||||||
Db     784 cattcgaacctattgggattttgaatttcctcgcccattcttaccaaatgttgattttgt 843

Qy     848 tggaggattccactggcaaacctgccaaacccctacctaaggaaatggaggagtttgtac 907
           |||||||  |||| | |||||  ||||||||||  |||||||||||||  |||||||| |
Db     844 tggaggacttcact-gtaaaccagccaaacccctgcctaaggaaatggaagagtttgtgc 902

Qy     908 agagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatga 967
           ||||||||||||||||||| ||||||||||||||||||||  |||| |||||||||
Db     903 agagctctggagaaaatggtattgtggtgttttctctggggtcgatgatcagtaacatgt 962

Qy     968 cagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgt 1027
           ||| |||||| |||||  | |||||| |||||||||| ||||||||||||||||||| |
Db     963 cagaagaaagtgccaacatgattgcatcagcccttgcccagatcccacaaaaggttctat 1022

Qy    1028 ggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataagtgga 1087
           |||||||||||| || || ||| || | ||||||  |||||||| ||||| ||||||
Db    1023 ggagatttgatggcaagaagccaaatactttaggttccaatactcgactgtacaagtggt 1082

Qy    1088 taccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtggag 1147
           |||||||||||||||||| |||||||| |||||||  |||||||||||||||||||||||
Db    1083 taccccagaatgaccttcttggtcatcccaaaaccaaagcttttataactcatggtggaa 1142

Qy    1148 ccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgtttt 1207
           |||||||||||||||||| ||||||||||||||||||||||||||||||||| ||||||
Db    1143 ccaatggcatctatgaggcgatctaccatgggatccctatggtgggcattcccttgtttg 1202

Qy    1208 gggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattggact 1267
            |||||||| |||||||||||||||||||| ||||||||||||||  | ||  |||||
Db    1203 cggatcaacatgataacattgctcacatgaaagccaagggagcagccctcagtgtggaca 1262

Qy    1268 tcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctt 1327
           |||  || ||||| |||| |||  |||||| |||| |||| ||||||| |||
Db    1263 tcaggaccatgtcaagtagagatttgctcaatgcattgaagtcagtcattaatgaccctg 1322

Qy    1328 tatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccc 1387
           | |||||||||||| | |||||||||||||||||||| ||||| |||||| | |||||||
Db    1323 tctataaagagaatgtcatgaaattatcaagaattcatcatgaccaaccaatgaagcccc 1382

Qy    1388 tggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttc 1447
           |||||||||||||||||||||||||||| |||||||||| ||||||||||||||| |||||||
Db    1383 tggatcgagcagtcttctggattgagtttgtcatgcgccacaaaggagccaagcaccttc 1442

Qy    1448 gagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttc 1507
           |||| ||||| ||  |||||||||||| ||||||||||||||||||||||||||  || |
Db    1443 gagtcgcagctcacaacctcacctggatccagtaccactctttggatgtgatagcattcc 1502

Qy    1508 tgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttct 1567
           |||||||||| |||||||||||||||||||||||||||||| ||||| ||||||||||||
Db    1503 tgctggcctgcgtggcaactgtgatatttatcatcacaaaattttgcctgttttgtttcc 1562

Qy    1568 ggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaag 1627
           | ||| |||| | || ||| |||||||  ||||||||||||||||| ||  |    |||||
Db    1563 gaaagcttgccaaaacaggaaagaagaagaaaagagattagttatatcaaaagcctgaag 1622

Qy    1628 ctggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtt 1687
            |||||   | || |||| ||||  || |  | | |       || | | |  | ||
Db    1623 -tggaatgactgaaagatgggactcctcctttattt------cagcatggagggttttaa 1675

Qy    1688 atgcaagatttctttcttcctgtgac 1713
           ||| | |||||| || ||||||||||
Db    1676 atggaggatttcctttttcctgtgac 1701


RESULT   6
Z95201
ID   Z95201 standard; DNA; 1686 BP.
XX
AC   Z95201;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B7 exon 1 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
```

XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 2; Page 45; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 1 of the human
CC UDP-glucuronosyltransferase 2B7 (UGT2B7) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,
CC also to predict their rate of metabolism of UGT2B substrate, potential
CC drug-drug interactions and adverse side effects. The polymorphisms can be
CC used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC linkage related to phenotypic variation in activity or expression of
CC UGT2B protein. The polymorphism containing nucleic acid molecules may
CC also be used for generating genetically modified non-human animals and
CC for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 1686 BP; 548 A; 286 C; 295 G; 557 T; 0 other;

Query Match 35.2%; Score 602.6; DB 21; Length 1686;
Best Local Similarity 89.2%; Pred. No. 2.3e-153;
Matches 661; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

```
Qy      1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
          || ||||||||||||||| || ||||||||||||||  || |||| ||||| ||  |
Db    390 attgcattgcaccaggatgtctgtgaaatggacttcagtaattttgctaatacaact-ga 448

Qy     61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
          | | || |||||||||||||| ||||||||| |||||||||||||| ||||||||||||
Db    449 gcttttgctttagctctgggaattgtggaaaggtgctggtgtgggcagcagaatacagcc 508

Qy    121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
          |||||||||||| |||||||||| | |||||| ||||||||||||||||||||||||
Db    509 attggatgaatataaagacaatcctggatgagcttattcagagaggtcatgaggtgactg 568

Qy    181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
          |||||||||||||||||||||||||||||||||||  |  |||| |||||||| |||
Db    569 tactggcatcttcagcttccattctttttgatcccaacaactcatccgctcttaaaattg 628

Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
          || ||||||| |||||||||||||||||||| || |||||| |||||||||||||| |||
Db    629 aaatttatcccacatctttaactaaaactgagttggagaatttcatcatgcaacagatta 688

Qy    301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
          |||||||||||||| ||| |||||||| |||||||||||||||||||||||| ||| ||||
Db    689 agagatggtcagaccttccaaaagatacatttggttatatttttcacaagtacaggaaa 748

Qy    361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
          || |||   |||  |||||||  |||||| ||||||||||||||||||||||||||||
Db    749 tcatgtcaatatttggtgacataactagaaagttctgtaaagatgtagtttcaaataaga 808

Qy    421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
          || |||||||||||| |||||||||||||||||||| || |||||||||||||| ||||||
Db    809 aatttatgaaaaaagtacaagagtcaagatttgacgtcatttttgcagatgctatttttc 868

Qy    481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
          ||||| |||||||||||||||| |||| ||||||||| |||||||||||||| ||||
Db    869 cctgtagtgagctgctggctgagctatttaacatacccttgtgtacagtctcagcttct 928

Qy    541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
          |||||||||||||  |||||| ||| |||||||||| | ||||||||||||||||| |||
Db    929 ctcctggctacacttttgaaaagcatagtggaggatttattttccctccttcctacgtac 988

Qy    601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
          || |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db    989 ctgttgttatgtcagaattaactgatcaaatgactttcatggagagggtaaaaaatatga 1048

Qy    661 tctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
          |||||||||||||| |||||||||||||| |||| | ||| ||||||||||||||||||
Db   1049 tctatgtgctttacttttgacttttggttcgaaatatttgacatgaagaagtgggatcagt 1108

Qy    721 tttacagtgaagttttaggaa 741
          |||| ||||||||| |||| |
Db   1109 tttatagtgaagttctaggta 1129
```

RESULT 7
Z95193
ID Z95193 standard; DNA; 1323 BP.
XX
AC Z95193;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UGT2B4 exon 1 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 1; Page 31; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 1 of the human
CC UDP-glucuronosyltransferase 2B4 (UGT2B4) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,

```
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 1323 BP; 403 A; 225 C; 253 G; 442 T; 0 other;


  Query Match             33.8%;  Score 579.2;  DB 21;  Length 1323;
  Best Local Similarity   85.7%;  Pred. No. 4.5e-147;
  Matches  655;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
           || |||||||| ||||||| || ||||||||||||||| |||||||||||||||| ||  |
Db     161 attgcattgcatcaggatgtctatgaaatggacttcagctcttctgctgatacagct-ga 219

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           | |||||||||||||||||||||||||||||| |||||||||||| || |||||| |||||
Db     220 gctgttactttagctctgggagttgtggaaaggtgctggtgtggcccacagaattcagcc 279

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           | |||||||||||| |||||||||||| | || ||||| |||||||||||||||||||||
Db     280 actggatgaatataaagacaatcctggatgaacttgtccagagaggtcatgaggtgactg 339

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           || ||||||||||||||||||  ||| ||||||||     |||| ||||||||||||
Db     340 tattggcatcttcagcttccatttctttcgatcccaacagcccatctactcttaaatttg 399

Qy     241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
           ||||||||||  |||||||||||||||||| |||||| |||| |||| ||| | |||||
Db     400 aagtttatcctgtatctttaactaaaactgagtttgaggatattatcaagcagctggtta 459

Qy     301 agagatggtcagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaa 360
           |||||||| ||||  ||| |||||| |  ||||||| |||||||||||||| ||||||||
Db     460 agagatgggcagaacttccaaaagacacattttggtcatatttttcacaagtacaagaaa 519

Qy     361 tcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataaga 420
           || |||||  |||  |||||||| |||||||| |||||||| ||| ||||||||||||||
Db     520 tcatgtggacatttaatgacatacttagaaagttctgtaaggatatagtttcaaataaga 579

Qy     421 aagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttc 480
           || ||||||| |||||||| |||||||||||||  | ||| ||||||||||||||| |
Db     580 aacttatgaagaaactacaggagtcaagatttgatgttgttcttgcagatgctgttttcc 639

Qy     481 cctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgcttta 540
           ||| |||||||||||||| | | ||||||| ||||  ||||| ||||| ||||||||
Db     640 cctttggtgagctgctggccgagttacttaaaatacccttttgtctacagcctccgcttct 699

Qy     541 ctcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacatac 600
           |||||||||| |||||||| ||| |||||||||  | ||||||||||||  | |
Db     700 ctcctggctacgcaattgaaaagcatagtggaggacttctgttccctccttcctatgtgc 759

Qy     601 ctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatga 660
           || |||||||||| || ||||||| |||||||||||||| ||||||||||||||||||||
Db     760 ctgttgttatgtcagaactaagtgaccaaatgactttcatagagagggtaaaaaatatga 819

Qy     661 tctatgtgctttatttgacttttggttccaaatgtctgatatgaagaagtgggatcagt 720
           |||||||||||||||||| |||||||||||||| | ||| ||||||||||||||||||
Db     820 tctatgtgctttatttgaattttggttccaaatatttgacatgaagaagtgggatcagt 879

Qy     721 tttacagtgaagttttaggaagacccactaccttatttgagaca 764
           | |||||||||| |||| |       |   | ||| |  |||
Db     880 tctacagtgaagttctaggtaagtaacttttttgattggtaaca 923
```

```
RESULT   8
Z95207
ID   Z95207 standard; DNA; 2312 BP.
XX
AC   Z95207;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B15 exon 1 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism;
KW   SNPs; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions
XX
PS   Example 3; Page 60-61; 72pp; English.
XX
CC   This sequence represents the nucleotide sequence of exon 1 of the human
CC   UDP-glucuronosyltransferase 2B15 (UGT2B15) gene.
CC   UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC   the glucuronic acid conjugation of a wide range of endogenous and
CC   exogenous substrates. The UGT2B gene subfamily encode steroid
CC   metabolizing isoforms in the liver. Alteration of the expression or
CC   function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC   can be used to detect altered UGT2B metabolism of a substrate in an
CC   individual. The nucleic acid molecules comprising a human UGT2B sequence
CC   polymorphism can be used in screening assays for genotyping individuals,
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 2312 BP; 765 A; 360 C; 414 G; 773 T; 0 other;


  Query Match             28.7%;  Score 491.4;  DB 21;  Length 2312;
  Best Local Similarity   79.2%;  Pred. No. 3.4e-123;
  Matches  609;  Conservative    0;  Mismatches  156;  Indels    4;  Gaps    2;

Qy       1 atcgcattgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctcca 60
           || ||||   |||||||| |||||||||||||| |||||  |||||||||||| || ||
Db     686 attgcataagaccaggatgtctctgaaatggacgtcagtctttctgctgatacagct-ca 744

Qy      61 gttgttactttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagcc 120
           |||||||||||||||||| || |||||||| ||||| |||||| || ||||||||||||
Db     745 gttgttactttagctctggaagctgtggaaaggtgctagtgtggcccacagaatacagcc 804

Qy     121 attggatgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactg 180
           ||||||| |||||||||||||||||| |||||||||||||| ||||||||||||||||
Db     805 attggataaatatgaagacaatcctggaagagcttgttcagaggggtcatgaggtgactg 864

Qy     181 tactggcatcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttg 240
           |  || ||||||| ||||| | |||| |  || ||| | |  ||||  || ||||||| |
Db     865 tgttgacatcttcggcttctactcttgtcaatgccagtaaatcatctgctattaaattag 924
```

```
Qy    241 aagtttatcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggtta 300
          ||||||||||||||||||||||||||||| ||| || ||  ||    |  || ||    |
Db    925 aagtttatcctacatctttaactaaaaatgatttggaagattctcttctgaaaattctcg 984

Qy    301 agagatgg---tcagacattcgaaaagatagcttttggttatatttttcacaagaacaag 357
          | ||||||   |  |   ||  |||| |||  ||||||| ||||||||||||  |||||
Db    985 atagatggatatatggtgtttcaaaaaatacattttggtcatatttttcacaattacaag 1044

Qy    358 aaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaata 417
          || |    |||||||  ||||||     ||  ||  ||||||||||| |||||  ||||
Db   1045 aattgtgttgggaatattatgactacagtaacaagctctgtaaagatgcagttttgaata 1104

Qy    418 agaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgttt 477
          ||||| ||||||  ||||||||||||||||  |||||  || || | ||||||||  ||
Db   1105 agaaacttatgatgaaactacaagagtcaaagtttgatgtcattctggcagatgcccttа 1164

Qy    478 ttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgct 537
          ||||||||||||||| |||||||  ||| |||||||||  ||| ||||||||||  || |
Db   1165 atccctgtggtgagctactggctgaactatttaacatacccttctgtacagtcttcgat 1224

Qy    538 ttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctaca 597
          |  ||  |||||||||| |||| | | |  ||||||||  |  | ||||||||||||||
Db   1225 tctctgttggctacacatttgagaagaatggtggaggatttctgttccctccttcctatg 1284

Qy    598 tacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaata 657
          ||||| |||||||||||| ||||||||||||||||||  |||||||||||||  |||||||||
Db   1285 tacctgttgttatgtcagaattaagtgatcaaatgattttcatggagaggataaaaaata 1344

Qy    658 tgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatc 717
          ||||  || ||||||||||||||||||||| ||||| | |||| ||||||||||||| |
Db   1345 tgatacatatgctttattttgacttttggtttcaaatttatgatctgaagaagtgggacc 1404

Qy    718 agttttacagtgaagttttaggaagacccactaccttatttgagacaat 766
          ||||||| |||||||| |||| |   |   |  || | | |   ||
Db   1405 agttttatagtgaagttctaggtaagtcatgtgtctaactggtgcttat 1453
```

RESULT 9
Z95205
ID Z95205 standard; DNA; 596 BP.
XX
AC Z95205;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UGT2B7 exon 5 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B7; UGT2B7; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 2; Page 47-48; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 5 of the human
CC UDP-glucuronosyltransferase 2B7 (UGT2B7) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,
CC also to predict their rate of metabolism of UGT2B substrate, potential
CC drug-drug interactions and adverse side effects. The polymorphisms can be
CC used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC linkage related to phenotypic variation in activity or expression of
CC UGT2B protein. The polymorphism containing nucleic acid molecules may
CC also be used for generating genetically modified non-human animals and
CC for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 596 BP; 199 A; 97 C; 115 G; 185 T; 0 other;

Query Match 19.1%; Score 326.6; DB 21; Length 596;
Best Local Similarity 92.5%; Pred. No. 8.2e-79;
Matches 356; Conservative 0; Mismatches 24; Indels 5; Gaps 1;

```
Qy   1329 atataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccct 1388
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     20 atataaagagaatgttatgaaattatcaagaattcaacatgatcaaccagtgaagcccct 79

Qy   1389 ggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcg 1448
          |||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db     80 ggatcgagcagtcttctggattgaatttgtcatgcgccacaaaggagctaaacaccttcg 139

Qy   1449 agttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttct 1508
            |||||||||||| ||||||||||||||||||||||||||||||||||||||||||| ||
Db    140 ggttgcagcccacgacctcacctggttccagtaccactctttggatgtgattgggttcct 199

Qy   1509 gctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1568
          ||||| |||||||||||||||||||||||| ||||||| | ||||||||||||||||||
Db    200 gctggtctgtgtggcaactgtgatatttatcgtcacaaaatgttgtctgttttgtttctg 259

Qy   1569 gaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagc 1628
          |||||||||||||||||  ||||||||||||  |||||||||| ||||| |||||||||
Db    260 gaagtttgctagaaaagcaaagaagggaaaaaatgattagttatatctgagatttgaagc 319

Qy   1629 tggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtta 1688
          ||||||||| ||||| |  ||| ||||||||||||||||||||     ||||||||| |
Db    320 tggaaaacctgataggtgagactacttcagtttattccagcaag-----aaagattgtga 374

Qy   1689 tgcaagatttctttcttcctgtgac 1713
          ||||||||||||||||||||| |||
Db    375 tgcaagatttctttcttcctgagac 399
```

RESULT 10
Q27370
ID Q27370 standard; cDNA; 2368 BP.
XX
AC Q27370;
XX
DT 27-JAN-1993 (first entry)
XX
DE HUG-Br2.
XX
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 30..801

```
FT                   /*tag= a
FT   polyA_signal    2347..2352
FT                   /*tag= b
FT                   /number= 1
FT   polyA_signal    2355..2360
FT                   /*tag= c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; R26154.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (Q27369) and HUGBr2
CC   (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2368 BP; 609 A; 532 C; 566 G; 661 T; 0 other;
```

Query Match 18.2%; Score 311.4; DB 13; Length 2368;
Best Local Similarity 53.0%; Pred. No. 2e-74;
Matches 789; Conservative 0; Mismatches 681; Indels 19; Gaps 5;

```
Qy     85 gtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaagacaatcc 144
Db    112 gtggaaaggtgttggtggtgcccactgatggcagcccctggctcagcatgcgggaggcct 171

Qy    145 tgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagcttccattc 204
Db    172 tgcgggagctccatgccagaggccaccaggcggtggtcctcaccccaagaggtgaatatg 231

Qy    205 tttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatctttaacta 264
Db    232 cacatcaagaagagaaatttttcaccctga------cagcctatgctgttccatggaccc 285

Qy    265 aaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacattcgaaaag 324
Db    286 agaaggaatttgatcgcgttacgctgggctacactcaagggt------tctttgaaacag 339

Qy    325 atagcttttggttatatttttcacaagaacaagaaatcctgtgggaattatatgacatat 384
Db    340 aacatcttctgaagagatattctagaagtatggcaattatgaacaatgtatctttggccc 399

Qy    385 ttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaactacaagagt 444
Db    400 ttcataggtgttgtgtggagctactgcataatgaggccctgatcaggcacctgaatgcta 459

Qy    445 caagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgctggctgcgc 504
Db    460 cttcctttgatgtggttttaacagaccccgttaacctctgcggggcggtgctggctaagt 519

Qy    505 tacttaacatacggtttgtgtacagtctccgctttactcctggctacacaattgaaaggc 564
Db    520 acctgtcgattcctgctgtgttttttggaggtacattccatgtgacttagactttaagg 579

Qy    565 acagtggaggactgattttccctccttcctacatacctattgttatgtcaaaattaagtg 624
Db    580 gcacacagtgtccaaat---ccttcctcctatattcctaagttactaacgaccaattcag 636

Qy    625 atcaaatgactttcatggagagggtaaaaaatatgatctatgtgctttatttgactttt 684
Db    637 accacatgacattcctgcaaagggtcaagaacatgctctaccctctggccctgtcctaca 696

Qy    685 ggttccaaatgtctgatatgaagaagtgggatcagttttacagtgaagttttaggaagac 744
Db    697 tttgccatacttttttctgccccttatgcaagtc---ttgcctctgagctttttcagagag 753

Qy    745 ccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaactcctgga 804
Db    754 aggtgactgtccaggacctattgagctctgcatctgtctggctgtttagaagtgactttg 813

Qy    805 gttttcaatttcctcatccattcttaccaaacgttgattttgttggaggattccactggc 864
Db    814 tgaaggattaccctaggcccatcatgcccaatatggtcttcattgggggcatcaactgtt 873

Qy    865 aaacctgccaaaccccctacctaaggaaatggaggagtttgtacagagctctggagaaaat 924
Db    874 gccaacgggaagccactatct-aggaatttgaagcctacattaatgcttctggagaacat 932

Qy    925 ggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagggccaat 984
Db    933 ggaattgtggttttctctttggaatcaatggtctcagaaattccagagaagaaagctatg 992

Qy    985 gtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgatgggaat 1044
Db    993 gcaattgctgatgctttgggcaaaatccctcagacagtcctgtggcggtacactggaacc 1052

Qy   1045 aaaccagatgccttaggtctcaatactcggctgtataagtggataccccagaatgacctt 1104
Db   1053 cgaccatcgaatcttgcgaacaacacgatacttgttaagtggctaccccaaaacgatctg 1112

Qy   1105 ctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatgag 1164
Db   1113 cttggtcacccgatgacccgtgcctttatcacccatgctggttcccatggtgtttatgaa 1172

Qy   1165 gcaatctaccatgggatccctatggtgggcattccattgttttgggatcaacctgataac 1224
Db   1173 agcatatgcaatggcgttcccatggtgatgatgcccttgtttggtgatcagatggacaat 1232

Qy   1225 attgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagt 1284
Db   1233 gcaaagcgcatggagactaagggagctggagtgaccctgaatgttctggaaatgacttct 1292

Qy   1285 acagacctgctgaatgcactgaagacagtaattaatgatccttttatataaagagaatatt 1344
Db   1293 gaagatttagaaaatgctcaaaaagcagtcatcaatgacaaaagttacaaggagaacatc 1352

Qy   1345 atgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgagcagtcttc 1404
Db   1353 atgcgcctctccagccttcacaaggaccgcccggtggagccgctggacctggccgtgttc 1412

Qy   1405 tggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcagcccatgac 1464
Db   1413 tgggtggagtttgtgatgaggcacaagggcgcgccacacctgcgccccgcagcccacgac 1472

Qy   1465 ctcacctggttccagtaccactctttggatgtgattgggtttctgctggcctgtgtggca 1524
Db   1473 ctcacctggtaccagtaccattccttggacgtgattggtttcctcttggccgtgctgctg 1532

Qy   1525 actgtgatatttatcatcacaaagttttgtctgttttgtttctggaagt 1573
Db   1533 acagtggccttcatcacctttaaatgttgtgcttatggctaccggaaat 1581
```

```
RESULT  11
Q27369
ID   Q27369 standard; cDNA; 2351 BP.
XX
AC   Q27369;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br1.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             16..784
FT                   /*tag=  a
FT   polyA_signal    2330..2335
FT                   /*tag=  b
FT                   /number= 1
FT   polyA_signal    2338..2343
FT                   /*tag=  c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; R26153.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (Q27369) and HUGBr2
CC   (Q27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;


  Query Match            17.2%;  Score 294.8;  DB 13;  Length 2351;
  Best Local Similarity  57.5%;  Pred. No. 6.1e-70;
  Matches  569;  Conservative    0;  Mismatches  417;  Indels    4;  Gaps    2;

Qy    584 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 643
          |||    |||||| | || |    | | ||         |||||  |||||  |||  || |
Db    579 cccattctcctacgtgcccaggcctctctcctctcattcagatcacatgaccttcctgca 638

Qy    644 gagggtaaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 703
          | |||| || || ||| ||  ||   |||      |||||  || | |  ||  | ||
Db    639 gcgggtgaagaacatgctcattgccttttcacagaactttctgtgcgacgtggtttattc 698

Qy    704 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 763
          | |       |   ||  |  ||| |  |    |||      ||  |    ||
Db    699 cccgtatgcaaccc---ttgcctcagaattccttcagagagaggtgactgtccaggacct 755

Qy    764 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 823
          | || |    ||     | |||||  |  |||    ||         | |  |||   ||
Db    756 attgagctctgcatctgtctggctgtttagaagtgactttgtgaaggattaccctaggcc 815

Qy    824 attcttaccaaacgttgattttgttggaggattccactggcaaacctgccaaacccctac 883
            || | || ||  | | |||||||||| ||| || |||| |    |    || || |||
Db    816 catcatgcccaatatggtttttgttggtggaatcaactgccttcaccaa-aatccactat 874

Qy    884 ctaaggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctc 943
          |  ||||| | || |  |   |  |    |||||||||| ||||  |||||||  || |||
Db    875 cccaggaatttgaagcctacattaatgcttctggagaacatggaattgtggttttctctt 934

Qy    944 tggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttg 1003
          |||  ||| || |    | ||  |||   | |  || |  | ||||||    ||  | |
Db    935 tggaatcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgg 994

Qy   1004 ccaagatcccacaaaaggttctgtggagatttgatgggaataaaccagatgccttaggtc 1063
           ||| ||||| || |  || |||||| | |    ||| |   ||||       | |
Db    995 gcaaaatccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcga 1054

Qy   1064 tcaatactcggctgtataagtggataccccagaatgaccttctaggtcatccaaaaacca 1123
           ||| ||    ||   ||||||| ||||||| || || || || ||||| || |  |||
Db    1055 acaacacgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgaccc 1114

Qy   1124 gagcttttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatcc 1183
          | || ||||| || |||| |||  || ||||  | |||||    || | |. ||||  | |
Db    1115 gtgcctttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttc 1174

Qy   1184 ctatggtgggcattccattgttttgggatcaacctgataacattgctcacatgaaggcca 1243
          | ||||||   || || |||||| | |||||    || ||       | |||| || | |
Db    1175 ccatggtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagacta 1234

Qy   1244 agggagcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcac 1303
          ||||||| |  || |   || |  |      |||| |   |  |||  |    ||||| |
Db    1235 agggagctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctc 1294

Qy   1304 tgaagacagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattc 1363
            ||  |||| || |||||       || || ||||| || |||_   | || ||  |||
Db    1295 aaaaagcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttc 1354

Qy   1364 aacatgatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgc 1423
          |  | || |  || ||  |||| ||||| |  || || |||||| | || ||||| |||
Db    1355 acaaggaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatga 1414

Qy   1424 cccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtacc 1483
            ||||| || ||   |||||| ||    |||||||| |||||||||||| ||||||||
Db    1415 ggcacaagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtacc 1474

Qy   1484 actctttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatca 1543
          | || ||||| |||||||| || ||  |||||    ||   || |||   || |||| |
Db    1475 attccttggacgtgattggtttcctcttggccgtgctgctgacagtggccttcatcacct 1534

Qy   1544 caaagttttgtctgttttgtttctggaagt 1573
            || | ||||   | | | | | |||| |
Db    1535 ttaaatgttgtgcttatggctaccggaaat 1564


RESULT  12
Z95198
ID   Z95198 standard; DNA; 1589 BP.
XX
AC   Z95198;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UGT2B4 exon 6 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
```

```
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M, Miller A, Penny L, Riedy M;
XX
DR   WPI; 2000-195321/17.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions -
XX
PS   Example 1; Page 33-34; 72pp; English.
XX
CC   This sequence represents the nucleotide sequence of exon 6 of the human
CC   UDP-glucuronosyltransferase 2B4 (UGT2B4) gene.
CC   UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC   the glucuronic acid conjugation of a wide range of endogenous and
CC   exogenous substrates. The UGT2B gene subfamily encode steroid
CC   metabolizing isoforms in the liver. Alteration of the expression or
CC   function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC   can be used to detect altered UGT2B metabolism of a substrate in an
CC   individual. The nucleic acid molecules comprising a human UGT2B sequence
CC   polymorphism can be used in screening assays for genotyping individuals,
CC   also to predict their rate of metabolism of UGT2B substrate, potential
CC   drug-drug interactions and adverse side effects. The polymorphisms can be
CC   used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC   linkage related to phenotypic variation in activity or expression of
CC   UGT2B protein. The polymorphism containing nucleic acid molecules may
CC   also be used for generating genetically modified non-human animals and
CC   for obtaining site specific gene modification in cell lines.
XX
SQ   Sequence 1589 BP; 467 A; 312 C; 293 G; 517 T; 0 other;

  Query Match             15.5%;  Score 266.2;  DB 21;  Length 1589;
  Best Local Similarity   86.0%;  Pred. No. 2.9e-62;
  Matches  331;  Conservative    0;  Mismatches   48;  Indels    6;  Gaps    3;

Qy   1329 atataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccct 1388
           ||||||||||||  |||||||||||||||||||| ||||||||||||| ||||||||
Db    731 atataaagagaatgctatgaaattatcaagaattcatcatgatcaaccagtgaagcccct 790

Qy   1389 ggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcg 1448
           |||||||||||||||||||||||||||||||| ||| |||||||||||| ||||||||
Db    791 tgatcgagcagtcttctggattgaatttgtcatgcgccataaaggagccaagcaccttcg 850

Qy   1449 agttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttct 1508
           |||||||||||| |||||||||||||||||||||||||||||||||||||| |||||| ||
Db    851 ggttgcagcccacgacctcacctggttccagtaccactctttggatgtgactgggttcct 910

Qy   1509 gctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1568
           |||||||||||||||||||||||||||| ||||||||||    |||||||||||| ||||
Db    911 gctggcctgtgtggcaactgtgatattcatcatcacaaa---atgtctgttttgtgtctg 967

Qy   1569 gaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagc 1628
           |||||||| ||||| ||| |||||||| |||||||||| ||| ||||||   | |||||
Db    968 gaagtttgttagaacaggaaagaaggggaaaagagattaattacgtctgaggctggaagc 1027

Qy   1629 tggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgtta 1688
           ||| |||||  ||| || | ||  ||| |||||||| || ||||||   |     || ||

Db   1028 tgggaaacccaataaat-gaactcctttagtttattacaacaagaa--gacgttgtgata 1084

Qy   1689 tgcaagatttctttcttcctgtgac 1713
                ||||| |||||||| ||||||
Db   1085 caagagattcctttcttcttgtgac 1109


RESULT 13
Z45118
ID   Z45118 standard; DNA; 735 BP.
XX
AC   Z45118;
XX
DT   28-FEB-2000 (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exons 2-5 nucleotide sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L, Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   P-PSDB; Y57100.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism -
XX
PS   Examples; Page 43-44; 63pp; English.
XX
CC   Nucleotide sequences Z45110-Z45118 are exons 1A-1J of the human uridine
CC   diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a family
CC   of enzymes that catalyse the glucuronic acid conjugation of a wide range
CC   of endogenous and exogenous substrates including phenols, alcohols,
CC   amines and fatty acids. Many of the reactions catalysed by UGTs result in
CC   toxic substances being converted to compounds which are more water
CC   soluble and are excreted. The invention relates to and identifies UGT1
CC   polymorphisms (Z45004-Z45041). The polymorphism sequences are useful as
CC   probes for detecting UGT1 locus polymorphisms, indicative of altered UGT1
CC   expression or activity. These polymorphisms are associated with
CC   Crigler-Najjar and Gilbert syndromes (unconjugated hyperbilirubinaemia)
CC   and drug metabolism. The genotyping of the UGT1 gene is used to predict
CC   the rate of metabolism of UGT1 substrates, possible drug-drug
CC   interactions and adverse side effects (i.e. to optimize drug dosage), and
CC   to screen for diseases caused by exposure to toxins and to study the
CC   effects of polymorphisms on enzymatic activity. The UGT1 sequences,
CC   including polymorphisms, can also be used to produce the corresponding
CC   protein (or its fragments) or to generate transgenic animals or modified
CC   cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence 735 BP; 189 A; 179 C; 187 G; 180 T; 0 other;

  Query Match             15.5%;  Score 265.8;  DB 21;  Length 735;
  Best Local Similarity   61.8%;  Pred. No. 2.6e-62;
  Matches  423;  Conservative    0;  Mismatches  262;  Indels    0;  Gaps    0;

Qy    889 gaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgggg 948
```

```
Db      1 gaatttgaagcctacattaatgcttctggagaacatggaattgtggtttttctctttggga 60

Qy    949 tcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaag 1008
Db     61 tcaatggtctcagaaattccagagaagaaagctatggcaattgctgatgctttgggcaaa 120

Qy   1009 atcccacaaaaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaat 1068
Db    121 atccctcagacagtcctgtggcggtacactggaacccgaccatcgaatcttgcgaacaac 180

Qy   1069 actcggctgtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagct 1128
Db    181 acgatacttgttaagtggctaccccaaaacgatctgcttggtcacccgatgacccgtgcc 240

Qy   1129 tttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatg 1188
Db    241 tttatcacccatgctggttcccatggtgtttatgaaagcatatgcaatggcgttcccatg 300

Qy   1189 gtgggcattccattgttttgggatcaacctgataacattgctcacatgaaggccaaggga 1248
Db    301 gtgatgatgcccttgtttggtgatcagatggacaatgcaaagcgcatggagactaaggga 360

Qy   1249 gcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaag 1308
Db    361 gctggagtgaccctgaatgttctggaaatgacttctgaagatttagaaaatgctctaaaa 420

Qy   1309 acagtaattaatgatccttttatataaagagaatattatgaaattatcaagaattcaacat 1368
Db    421 gcagtcatcaatgacaaaagttacaaggagaacatcatgcgcctctccagccttcacaag 480

Qy   1369 gatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccac 1428
Db    481 gaccgcccggtggagccgctggacctggccgtgttctgggtggagtttgtgatgaggcac 540

Qy   1429 aaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactct 1488
Db    541 aagggcgcgccacacctgcgccccgcagcccacgacctcacctggtaccagtaccattcc 600

Qy   1489 ttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaag 1548
Db    601 ttggacgtgattggtttcctcttggccgtcgtgctgacagtggccttcatcacctttaaa 660

Qy   1549 ttttgtctgttttgtttctggaagt 1573
Db    661 tgttgtgcttatggctaccggaaat 685
```

RESULT 14
Z95211
ID Z95211 standard; DNA; 978 BP.
XX
AC Z95211;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UGT2B15 exon 5 nucleotide sequence.
XX
KW UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW drug interaction; detect; human; single nucleotide polymorphism;
KW SNPs; ds.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Example 3; Page 62; 72pp; English.
XX
CC This sequence represents the nucleotide sequence of exon 5 of the human
CC UDP-glucuronosyltransferase 2B15 (UGT2B15) gene.
CC UDP-glucuronosyltransferase (UGTs) are a family of enzymes that catalyse
CC the glucuronic acid conjugation of a wide range of endogenous and
CC exogenous substrates. The UGT2B gene subfamily encode steroid
CC metabolizing isoforms in the liver. Alteration of the expression or
CC function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms. Probes which detect the UGT2B locus polymorphisms
CC can be used to detect altered UGT2B metabolism of a substrate in an
CC individual. The nucleic acid molecules comprising a human UGT2B sequence
CC polymorphism can be used in screening assays for genotyping individuals,
CC also to predict their rate of metabolism of UGT2B substrate, potential
CC drug-drug interactions and adverse side effects. The polymorphisms can be
CC used as single nucleotide polymorphisms (SNPs) for detecting genetic
CC linkage related to phenotypic variation in activity or expression of
CC UGT2B protein. The polymorphism containing nucleic acid molecules may
CC also be used for generating genetically modified non-human animals and
CC for obtaining site specific gene modification in cell lines.
XX
SQ Sequence 978 BP; 321 A; 187 C; 162 G; 308 T; 0 other;

Query Match 14.4%; Score 246; DB 21; Length 978;
Best Local Similarity 81.2%; Pred. No. 6.9e-57;
Matches 312; Conservative 0; Mismatches 65; Indels 7; Gaps 2;

```
Qy   1330 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagccccctg 1389
Db    327 tataaagagaatgtcatgaaattatcaagaattcatcatgaccaaccaatgaagccccctg 386

Qy   1390 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1449
Db    387 gatcgagcagtcttctggattgagtttgtcatgcgccacaaaggagccaagcaccttcga 446

Qy   1450 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1509
Db    447 gtcgcagctcacaacctcacctggatccagtaccactctttggatgtgatagcattcctg 506

Qy   1510 ctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctgg 1569
Db    507 ctggcctgcgtggcaactgtgatatttatcatcacaaaattttgcctgttttgtttccga 566

Qy   1570 aagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttgaagct 1629
Db    567 aagcttgccaaaacaggaaagaagaagaaaagagattagttatatcaaaagcctgaag-t 625

Qy   1630 ggaaaaccagatagataggacaacttcagtttattccagcaagaaagaaaagattgttat 1689
Db    626 ggaatgactgaaagatgggactcctcctttattt------cagcatggagggttttaaat 679

Qy   1690 gcaagatttctttcttcctgtgac 1713
Db    680 ggaggatttccttttcctgtgac 703
```

RESULT 15
C03286
ID C03286 standard; cDNA; 350 BP.
XX
AC C03286;
XX

DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3284.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; G03280.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 3284; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 350 BP; 108 A; 69 C; 77 G; 96 T; 0 other;

Query Match 12.9%; Score 220.6; DB 21; Length 350;
Best Local Similarity 78.6%; Pred. No. 3.3e-50;
Matches 276; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

```
Qy    777 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 836
          || |||||||| || ||||  |  |||  |||| |||||||||| |||| |  ||| |||
Db      1 gagatatggctaatacgaacatattgggattttgaatttcctcaaccataccaacctaac 60

Qy    837 gttgattttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatgga 896
          |||| |||||||||||||| |||| | ||||||||||||  |  | ||||||||||||||
Db     61 tttgagtttgttggaggattgcact-gtaaacctgccaaagctttgcctaaggaaatgga 119

Qy    897 ggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgat 956
           | ||||| ||||| || || ||| ||||| ||||||||||||||||||||||| || |
Db    120 aaattttgtccagagttcagggaagatggtattgtggtgttttctctggggtcactgtt 179

Qy    957 aagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccaca 1016
            ||  | |||| ||||| ||| ||| | |||||  |||||||||| ||||||||||
Db    180 tcaaaatgttacagaagaaaaggctaatatcattgcttcagcccttgcccagatcccaca 239

Qy   1017 aaaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggct 1076
          |||||  | ||||| |   | || || ||||||    | |||||   ||||||||||
Db    240 gaaggtgttatggaggtacaaaggaaaaaaaccatccacattaggagccaatactcggct 299

Qy   1077 gtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagc 1127
          |||| | |||||||||||||||| ||||| |||||||| ||||||| |||
Db    300 gtatgattggataccccagaatgatcttcttggtcatcccaaaaccaaagc 350
```

Search completed: April 30, 2001, 19:04:50
Job time: 36704 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:31:10 ; Search time 330.07 Seconds
(without alignments)
906.154 Million cell updates/sec

Title: US-09-721-183-2
Perfect score: 1713
Sequence: 1 atcgcattgcaccaggatga.....gatttctttcttcctgtgac 1713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	319.4	18.6	2339	5	PCT-US92-00282-2	Sequence 2, Appli
	2	298	17.4	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	3	106.6	6.2	1561	5	PCT-US92-00282-25	Sequence 25, Appl
	4	81.6	4.8	783	5	PCT-US92-00282-22	Sequence 22, Appl
	5	59.2	3.5	1008	5	PCT-US92-00282-10	Sequence 10, Appl
	6	57.6	3.4	1066	5	PCT-US92-00282-14	Sequence 14, Appl
	7	52.8	3.1	1197	5	PCT-US92-00282-12	Sequence 12, Appl
	8	52.8	3.1	1448	5	PCT-US92-00282-16	Sequence 16, Appl
	9	49.6	2.9	513	5	PCT-US92-00282-20	Sequence 20, Appl
	10	46.2	2.7	51	5	PCT-US92-00282-37	Sequence 37, Appl
	11	42.2	2.5	1669	2	US-08-522-421-1	Sequence 1, Appli
	12	42.2	2.5	1800	6	5180581-1	Patent No. 5180581
	13	42.2	2.5	2793	1	US-08-281-916-5	Sequence 5, Appli
	14	42.2	2.5	2793	2	US-08-460-725-7	Sequence 7, Appli
	15	39.6	2.3	1190	5	PCT-US92-00282-18	Sequence 18, Appl
	16	37.2	2.2	22846	2	US-08-469-461-3	Sequence 3, Appli
	17	37.2	2.2	22846	3	US-07-890-609-3	Sequence 3, Appli
	18	37	2.2	2781	3	US-08-749-522-4	Sequence 4, Appli
c	19	36.8	2.1	6152	4	US-08-973-462-1	Sequence 1, Appli
	20	36.6	2.1	1618	1	US-07-885-970A-24	Sequence 24, Appl
	21	36.6	2.1	1618	1	US-08-298-687A-24	Sequence 24, Appl
	22	36.6	2.1	1618	1	US-08-530-797-14	Sequence 14, Appl
	23	36.6	2.1	1618	1	US-08-298-829-24	Sequence 24, Appl
	24	36.6	2.1	1618	2	US-08-787-335-14	Sequence 14, Appl
	25	36.6	2.1	1627	3	US-09-106-464-1	Sequence 1, Appli
	26	36.4	2.1	1607	2	US-08-797-226-1	Sequence 1, Appli
	27	36	2.1	1219	5	PCT-US92-00282-8	Sequence 8, Appli
c	28	35	2.0	2749	1	US-08-344-695-1	Sequence 1, Appli
c	29	34.8	2.0	1075	4	US-08-960-780-30	Sequence 30, Appl
c	30	34.8	2.0	3095	6	5231168-1	Patent No. 5231168
	31	34.8	2.0	3336	4	US-09-330-970-4	Sequence 4, Appli
	32	34.6	2.0	790	6	5194596-8	Patent No. 5194596
c	33	34.4	2.0	669	2	US-09-109-266-7	Sequence 7, Appli
c	34	34	2.0	1183	2	US-08-731-722-8	Sequence 8, Appli
c	35	34	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	36	33.8	2.0	1785	1	US-08-592-696-3	Sequence 3, Appli
	37	33.8	2.0	1785	2	US-09-027-536-3	Sequence 3, Appli
	38	33.8	2.0	1785	3	US-09-028-148-3	Sequence 3, Appli
	39	33.8	2.0	2265	1	US-08-242-689-1	Sequence 1, Appli
c	40	33.4	1.9	1772	3	US-08-793-666-8	Sequence 8, Appli
c	41	33.4	1.9	1772	3	US-08-793-666-11	Sequence 11, Appl
	42	33.4	1.9	3781	3	US-08-688-988-5	Sequence 5, Appli
c	43	33	1.9	5232	4	US-09-212-971-3	Sequence 3, Appli
c	44	33	1.9	5232	4	US-08-800-929A-3	Sequence 3, Appli
	45	32.8	1.9	654	1	US-08-257-073-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT   1
PCT-US92-00282-2
; Sequence 2, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2339 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-2


  Query Match           18.6%;  Score 319.4;  DB 5;  Length 2339;
  Best Local Similarity  53.3%;  Pred. No. 8.6e-81;
  Matches  794;  Conservative    0;  Mismatches  676;  Indels   19;  Gaps    5;

Qy      85 gtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaagacaatcc 144
           ||||||| ||| |||||  | || | ||   ||||| ||| | |  |||  |     |
Db      83 GTGGAAAGGTGTTGGTGGTGCCCACTGATGGCAGCCCCTGGCTCAGCATGCGGGAGGCCT 142
```

```
Qy    145 tgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagcttccattc 204
          ||   |||||   |   |||||  ||  |||  |   ||  ||  |  |     |      |
Db    143 TGCGGGAGCTCCATGCCAGAGGCCACCAGGCGGTGGTCCTCACCCCAGAGG------TGA 196

Qy    205 tttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatctttaactа 264
          | |  |   ||| ||  |       | | |   |   ||  ||| ||    | |  ||
Db    197 ATATGCACATCAAAGAAGAGAAATTTTTCACCCTGACAGCCTATGCTGTTCCATGGACCC 256

Qy    265 aaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacattcgaaaag 324
          | |  ||||||||     | |   ||    |   | |  | |       | ||  || ||
Db    257 AGAAGGAATTTGATCGCGTTACGCTGGGCTACACTCAAGGGT------TCTTTGAAACAG 310

Qy    325 atagcttttggttatatttttcacaagaacaagaaatcctgtgggaattatatgacatat 384
          |     ||  |      | |||   |      | |||  ||    |  ||| |
Db    311 AACATCTTCTGAAGAGATATTCTAGAAGTATGGCAATTATGAACAATGTATCTTTGGCCC 370

Qy    385 ttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaactacaagagt 444
          ||   |  |  |||   ||  || |    ||| ||    | ||  |  |  ||  | |
Db    371 TTCATAGGTGTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGCACCCTGAATGCTA 430

Qy    445 caagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgctggctgcgc 504
          |    |||||  | |||||  ||||  | |||    | |||  || | | ||||||||  |
Db    431 CTTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGGGGCGGTGCTGGCTAAGT 490

Qy    505 tacttaacatacggtttgtgtacagtctccgctttactcctggctacacaattgaaaggc 564
            ||    || |    |||||    |    | |  | |||  |  ||  |      | |
Db    491 ACCTGTCGATTCCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGACTTAGACTTTAAGG 550

Qy    565 acagtggaggactgattttccctccttcctacatacctattgttatgtcaaaattaagtg 624
          ||      | |  | |   ||| | ||||| || ||||   |  |  | |      |
Db    551 GCACACAGTGTCCAAAT---CCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAG 607

Qy    625 atcaaatgactttcatggagagggtaaaaaatatgatctatgtgctttattttgacttttt 684
          | || ||||| ||| || | ||||| || || ||| ||||    ||    |  ||
Db    608 ACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTACCCTCTGGCCCTGTCCTACA 667

Qy    685 ggttccaaatgtctgatatgaagaagtgggatcagttttacagtgaagttttaggaagac 744
          | ||| |  | |  |       |      ||  ||  |  |||  ||||    |||
Db    668 TTTGCCATACTTTTTCTGCCCCTTATGCAAGTC---TTGCCTCTGAGCTTTTTCAGAGAG 724

Qy    745 ccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaactcctgga 804
            |  |  | ||      |  |  | ||   | | |||||  | |||      ||
Db    725 AGGTGTCAGTGGTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTG 784

Qy    805 gttttcaatttcctcatccattcttaccaaacgttgattttgttggaggattccactggc 864
          |  | |  ||    ||  || | ||  ||| | |  ||  |||| ||  || ||||
Db    785 TGATGGACTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTG 844

Qy    865 aaacctgccaaaccccтacctaaggaaatggaggagtttgtacagagctctggagaaaat 924
          | | |  || || ||| || ||||| | || |   |   |  |    |||||||| ||
Db    845 CCAAC-GGGAAGCCACTATCTCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACAT 903

Qy    925 ggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagggccaat 984
          ||  ||||||| || ||| |||| ||| || |      | ||  |||   | |  || |
Db    904 GGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATG 963

Qy    985 gtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgatgggaat 1044
          | ||||||    ||  | | ||| ||||| || |  || ||||||  | |    ||| |
Db    964 GCAATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACC 1023

Qy   1045 aaaccagatgccttaggtctcaatactcggctgtataagtggataccccagaatgaccttt 1104
          ||||       | |    ||| ||    ||   ||||||| ||||||| || || ||
Db   1024 CGACCATCGAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTG 1083

Qy   1105 ctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatgag 1164
          || ||||| || |  ||| | || ||||| || |||| |||  || ||||  | |||||
Db   1084 CTTGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAA 1143

Qy   1165 gcaatctaccatgggatccctatggtgggcattccattgttttgggatcaacctgataac 1224
          ||  | |  |||| | || ||||||  || || ||||||  | |||||    || ||
Db   1144 AGCATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAAT 1203

Qy   1225 attgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagt 1284
               | |||| || | ||||||||  |  || |   || |  |      ||||  |   |
Db   1204 GCAAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCT 1263

Qy   1285 acagacctgctgaatgcactgaagacagtaattaatgatccтttatataaagagaatatt 1344
            |||  |    ||||| || ||  |||| || |||||  -     || || |||||  ||
Db   1264 GAAGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATC 1323

Qy   1345 atgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgagcagtcttc 1404
          | |    | || ||  |||| | || |  || ||  |||| |||||  ||    |||  |||
Db   1324 AGGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTC 1383

Qy   1405 tggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcagcccatgac 1464
          ||| | || ||||| |||   ||||| || ||   ||||||  ||    |||||||| |||
Db   1384 TGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGAC 1443

Qy   1465 ctcacctggttccagtaccactctttggatgtgattgggtttctgctggcctgtgtggca 1524
          |||||||||| |||||||||  || ||||| |||||||  || ||  |||||   |||
Db   1444 CTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTG 1503

Qy   1525 actgtgatatttatcatcacaaagttttgtctgttttgtttctggaagt 1573
          || ||| . || |||| |   || | |||   | | | | | |||| |
Db   1504 ACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAAT 1552


RESULT   2
PCT-US92-00282-1
; Sequence 1, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2336 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-1


  Query Match            17.4%;  Score 298;  DB 5;  Length 2336;
  Best Local Similarity  57.7%;  Pred. No. 9.8e-75;
  Matches  571;  Conservative    0;  Mismatches  415;  Indels    4;  Gaps    2;

Qy    584 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 643
```

```
Db    564 CCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACATGACCTTCCTGCA 623

Qy    644 gagggtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 703
Db    624 GCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGCGACGTGGTTTATTC 683

Qy    704 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 763
Db    684 CCCGTATGCAACCC---TTGCCTCAGAATTCCTTCAGAGAGAGGTGACTGTCCAGGACCT 740

Qy    764 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 823
Db    741 ATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGATTACCCTAGGCC 800

Qy    824 attcttaccaaacgttgattttgttggaggattccactggcaaacctgccaaacccctac 883
Db    801 CATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCAA-AATCCACTAT 859

Qy    884 ctaaggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctc 943
Db    860 CCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTT 919

Qy    944 tggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttg 1003
Db    920 TGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGG 979

Qy   1004 ccaagatcccacaaaaggttctgtggagatttgatgggaataaaccagatgccttaggtc 1063
Db    980 GCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGA 1039

Qy   1064 tcaatactcggctgtataagtggataccccagaatgaccttctaggtcatccaaaaacca 1123
Db   1040 ACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCC 1099

Qy   1124 gagcttttataactcatggtggagccaatggcatctatgaggcaatctaccatgggatcc 1183
Db   1100 GTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTC 1159

Qy   1184 ctatggtgggcattccattgttttgggatcaacctgataacattgctcacatgaaggcca 1243
Db   1160 CCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTA 1219

Qy   1244 agggagcagctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcac 1303
Db   1220 AGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTC 1279

Qy   1304 tgaagacagtaattaatgatcctttatataaagagaatattatgaaattatcaagaattc 1363
Db   1280 TAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCTCTCCAGCCTTC 1339

Qy   1364 aacatgatcaaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgc 1423
Db   1340 ACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGA 1399

Qy   1424 cccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtacc 1483
Db   1400 GGCACAAGGGCGCCCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACC 1459

Qy   1484 actctttggatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatca 1543
Db   1460 ATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCT 1519

Qy   1544 caaagttttgtctgttttgtttctggaagt 1573
Db   1520 TTAAATGTTGTGCTTATGGCTACCGGAAAT 1549


RESULT   3
PCT-US92-00282-25
; Sequence 25, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1561 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-25


  Query Match             6.2%;  Score 106.6;  DB 5;  Length 1561;
  Best Local Similarity   67.3%;  Pred. No. 1.1e-20;
  Matches  165;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy   1330 tataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctg 1389
Db    272 TACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTG 331

Qy   1390 gatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcga 1449
Db    332 GACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGC 391

Qy   1450 gttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1509
Db    392 CCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTC 451

Qy   1510 ctggcc-tgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctg 1568
Db    452 TTGGCCGTCTGGCTGGACAGTGGCCTTCATCACCTTTAAATGTTGTGCTTATGGCTACCG 511

Qy   1569 gaagt 1573
Db    512 GAAAT 516


RESULT   4
PCT-US92-00282-22
; Sequence 22, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  783 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-22


  Query Match             4.8%;  Score 81.6;  DB 5;  Length 783;
  Best Local Similarity   61.1%;  Pred. No. 9.5e-14;
  Matches  132;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy   1107 aggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctatgaggc 1166
          |||||| || |  ||| | || ||||| || |||| |||  || ||||  | |||||
Db    470 AGGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAG 529

Qy   1167 aatctaccatgggatccctatggtgggcattccattgttttgggatcaacctgataacat 1226
           || | | ||||  | || ||||||   || || |||||| | |||||    || ||
Db    530 CATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGC 589

Qy   1227 tgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagtac 1286
               | |||| || | |||||||| |  || |   || |  |       |||| |   |
Db    590 AAAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGA 649

Qy   1287 agacctgctgaatgcactgaagacagtaattaatga 1322
          |||  |    ||||| || ||  |||| || |||||
Db    650 AGATTTAGAAAATGCTCTAAAAGCAGTCATCAATGA 685


RESULT   5
PCT-US92-00282-10
; Sequence 10, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-10


  Query Match             3.5%;  Score 59.2;  DB 5;  Length 1008;
  Best Local Similarity   47.8%;  Pred. No. 2.3e-07;
  Matches  277;  Conservative    0;  Mismatches  288;  Indels   15;  Gaps    3;

Qy     85 gtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaagacaatcc 144
          |||| || |||||||||  | || | ||   |||||| ||| | |  |||  |     |
Db    194 GTGGGAAGGTGCTGGTGGTGCCCACTGATGGCAGCCACTGGCTCAGCATGCGGGAGGCCT 253

Qy    145 tgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagcttccattc 204
          ||   || ||   |  |||||| ||  |||||   || ||  |       |      |
Db    254 TGCGGGACCTCCATGCGAGAGGCCACCAGGTGGTGGTCCTCACCCTGGAGG------TGA 307

Qy    205 tttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatctttaacta 264
          | |  |   ||| || |    |  | | |   |   |   ||| | |  || |  ||
Db    308 ATATGTACATCAAAGAAGAGAACTTTTTCACCCTGACAACGTATGCCATTTCATGGACCC 367

Qy    265 aaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacattcgaaaag 324
          |    ||||||||         | |    | || | | |     ||   | ||  || ||
Db    368 AGGACGAATTTGA------TCGCCTTTTGCTGGGTCACACTCAATCGTTCTTTGAAACAG 421

Qy    325 atagcttttggttatatttttcacaagaacaagaaatcctgtgggaattatatgacatat 384
          |     ||  | |  | |||||   | |   | |||  ||    |  | | |   ||
Db    422 AACATCTTCTGATGAAATTTTCTAGAAGAATGGCAATTATGAACAATATGTCTTTGATCA 481

Qy    385 ttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaactacaagagt 444
          |    |  |  |||   ||  || |    ||| ||    | || |   | || || |
Db    482 TACATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGCATGCTA 541

Qy    445 caagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgctggctgcgc 504
          |    |||||  | ||| |  ||||  |  ||   | ||| |  | | ||||||||  |
Db    542 CTTCCTTTGATGTGGTTCTAACAGACCCCTTTCACCTCTGCGCGGCGGTGCTGGCTAAGT 601

Qy    505 tacttaacatacggtttgtgtacagtctccgctttactcctggctacacaattgaaaggc 564
            ||    || |    |||||       |  |    | |||  |  |   |      | |
Db    602 ACCTGTCGATTCCTGCTGTGTTTTTCTTGAGGAACATTCCATGTGATTTAGACTTTAAGG 661

Qy    565 acagtggaggactgattttccctccttcctacatacctattgttatgtcaaaattaagtg 624
          ||      |  |   |||| | ||||| || ||||   |  |  | |        |
Db    662 GCACACAGTGTCCAA---ACCCTTCCTCCTATATTCCTAGATTACTAACGACCAATTCAG 718

Qy    625 atcaaatgactttcatggagagggtaaaaaatatgatcta 664
          | || ||||| ||| || | |||||  || || ||| ||||
Db    719 ACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTA 758


RESULT   6
PCT-US92-00282-14
; Sequence 14, Application PC/TUS9200282
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-14


  Query Match             3.4%;  Score 57.6;  DB 5;  Length 1066;
  Best Local Similarity   47.6%;  Pred. No. 6.8e-07;
  Matches  276;  Conservative    0;  Mismatches  289;  Indels   15;  Gaps    3;

Qy      85 gtggaaaagtgctggtgtgggccgcagaatacagccattggatgaatatgaagacaatcc 144
           ||||||| ||| |||||  | ||    |   |||||| ||| | |  |||  |    ||
Db     213 GTGGAAAGGTGTTGGTGGTGCCCATTCATGGCAGCCACTGGCTCAGCATGCGGGAGGTCT 272

Qy     145 tgaaagagcttgttcagagaggtcatgaggtgactgtactggcatcttcagcttccattc 204
           ||   |||||    |   ||||| ||  |||     || ||  |  |     |      |
Db     273 TGCGGGAGCTCCATGCCAGAGGCCACCAGGCAGTGGTCCTCACCCCAGAGG------TGA 326

Qy     205 tttttgatcccaatgatgcatccactcttaaatttgaagtttatcctacatctttaacta 264
           | |  |   ||| || |    |  | | |    |   |   ||| | |  || |  ||
Db     327 ATATGCACATCAAAGAAGAGAACTTTTTCACCCTGACAACCTATGCCATTTCGTGGACCC 386

Qy     265 aaactgaatttgagaatatcatcatgcaacaggttaagagatggtcagacattcgaaaag 324
           |   |||||||||         |  ||   ||   | |     |       || ||  || ||
Db     387 AGGATGAATTTGATCGCCATGTGCTGGGCCACACTCAACTGT------ACTTTGAAACAG 440

Qy     325 atagcttttggttatatttttcacaagaacaagaaatcctgtgggaattatatgacatat 384
           |    |||  |    | ||||    |       | |||  ||    |  | | |    | |
Db     441 AACATTTTCTGAAGAAATTTTTCAGAAGTATGGCAATGTTGAACAATATGTCTTTGGTCT 500

Qy     385 ttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaaaactacaagagt 444
           |   |  |  |||   ||  || |    ||| ||    | || |   | ||  | |
Db     501 ATCATAGGTCTTGTGTGGAGCTACTACATAATGAGGCCCTGATCAGGCACCTGAATGCTA 560

Qy     445 caagatttgacatcgtttttgcagatgctgtttttccctgtggtgagctgctggctgcgc 504
           |    |||||  | |||||  ||||  | |||   | ||| |  |   ||||||||  |
Db     561 CTTCCTTTGATGTGGTTTTAACAGACCCCGTTAACCTCTGCGCGGCAGTGCTGGCTAAGT 620

Qy     505 tacttaacatacggtttgtgtacagtctccgctttactcctggctacacaattgaaaggc 564
             ||    || |    |||||    | |  |    | |||  |  |   |      | |

Db     621 ACCTGTCGATTCCTACTGTGTTTTTTTTGAGGAACATTCCATGTGATTTAGACTTTAAGG 680

Qy     565 acagtggaggactgattttccctccttcctacatacctattgttatgtcaaaattaagtg 624
             ||      | |  |    |||| | ||||| || ||||   |  |  |||       |
Db     681 GCACACAGTGTCCAA---ACCCTTCCTCCTATATTCCTAGATTACTAACAACCAATTCAG 737

Qy     625 atcaaatgactttcatggagagggtaaaaaatatgatcta 664
           | || ||||| |||||| | ||||| || || ||| ||||
Db     738 ACCACATGACATTCATGCAAAGGGTCAAGAACATGCTCTA 777


RESULT   7
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1197 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-12


  Query Match             3.1%;  Score 52.8;  DB 5;  Length 1197;
  Best Local Similarity   46.8%;  Pred. No. 1.7e-05;
  Matches  237;  Conservative    0;  Mismatches  262;  Indels    7;  Gaps    2;

Qy     357 gaaatcctgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaat 416
           | |||  ||    |  ||| |      ||   |  | |||   ||  || |    |||
Db     573 GCAATTATGAACAATGTATCTTTGGCCCTTCATAGGTGTTGTGTGGAGCTACTGCATAAT 632

Qy     417 aagaaagttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtt 476
             ||    | || |   | ||  | |   |    |||||  | |||||  ||||  | |||
Db     633 GAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCCGTT 692

Qy     477 tttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgc 536
             | |||||| | | ||||||||  |   ||    || |   |||||    |    |
Db     693 AACCTCTGTGGGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTTTTTTGGAGG 752

Qy     537 tttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctac 596
           |  | |||  |  ||  |     | |  ||      | |  ||    ||| | |||||
Db     753 TACATTCCATGTGACTTAGACTTTAAGGCCATACAGTGTCAAAT----CCTTCCTCCTAT 808
```

```
Qy    597 atacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaat 656
          || ||||   |  |  | |        || || ||||| ||| || | ||||| || ||
Db    809 ATTCCTAAGTTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAAC 868

Qy    657 atgatctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggat 716
          ||| ||||   ||     |  ||     | ||| |  | | |       |      |
Db    869 ATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCTTATGCAAGT 928

Qy    717 cagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagct 776
          |  ||  |  |||  ||||   |||       |  | | ||     |  |  | ||
Db    929 C---TTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTCAGCTATGCA 985

Qy    777 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 836
            | | |||||  | |||     ||      |  | |  ||    ||  || | || |||
Db    986 TCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCGATCATGCCCAAC 1045

Qy    837 gttgattttgttggaggattccactg 862
            | |  ||  |||| ||  || ||||
Db   1046 ATGGTCTTCATTGGGGGCATCAACTG 1071
```

```
RESULT   8
PCT-US92-00282-16
; Sequence 16, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1448 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
PCT-US92-00282-16
```

```
  Query Match             3.1%;  Score 52.8;  DB 5;  Length 1448;
  Best Local Similarity   48.9%;  Pred. No. 1.8e-05;
  Matches  172;  Conservative    0;  Mismatches  177;  Indels    3;  Gaps    1;

Qy    313 acattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtgggaat 372
          || ||  || |||    |||  |      |||||  ||      | |||  ||    | |
Db    404 ACTTTGAAACAGAACATTTTCTGAAGACATTTTCTAAAAGTATGGCAATTTTGAAAAATT 463
```

```
Qy    373 tatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatgaaaa 432
          || |    | |||   |  |  |||   ||  || |    ||| ||    | || |
Db    464 CATCTTTGGTCTTTCATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGC 523

Qy    433 aactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggtgagc 492
          | ||  | |   |    || ||  | |||||   |||  |  |||  | |||||| |
Db    524 ACCTGAATGCTACTTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCTGTGGGGCAG 583

Qy    493 tgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggctaca 552
          ||||||||  |   ||    || |    |||||    | |  |    | |||  | |
Db    584 TGCTGGCTAAGTACCTATCAATTCCTGCTGTGTTTTTTTTGAGGAACATTCCATGTGATT 643

Qy    553 caattgaaaggcacagtggaggactgattttccctccttcctacatacctattgttatgt 612
          |       | |  ||       | |  | |   ||| | ||||| || ||||    |  |
Db    644 TAGACTTTAAGGGCACACAGTGTCCAAAT---CCTTCCTCCTATATTCCTAAGTTACTAA 700

Qy    613 caaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatcta 664
          | |         || || ||||| ||| || | ||||| || || ||| ||||
Db    701 CGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAGAACATGCTCTA 752
```

```
RESULT   9
PCT-US92-00282-20
; Sequence 20, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-20
```

```
  Query Match             2.9%;  Score 49.6;  DB 5;  Length 513;
  Best Local Similarity   60.3%;  Pred. No. 8.8e-05;
  Matches   82;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy    887 aggaaatggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctgg 946
          ||||| | || |  |   |  |    ||||||||| ||||  ||||||| || ||| |||
Db    229 AGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGG 288
```

```
Qy    947 ggtcagtgataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgcca 1006
          | ||| || |     | ||  |||   | |  || |  | ||||||    ||  | | ||
Db    289 GATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCA 348

Qy   1007 agatcccacaaaaggt 1022
          | ||||| || |  ||
Db    349 AAATCCCTCAGACAGT 364


RESULT 10
PCT-US92-00282-37
; Sequence 37, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-37


  Query Match             2.7%;  Score 46.2;  DB 5;  Length 51;
  Best Local Similarity   94.1%;  Pred. No. 0.00026;
  Matches   48;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   1459 catgacctcacctggttccagtaccactctttggatgtgattgggtttctg 1509
          || ||||||||||||||||||||||||||||||||||||||||||||| || |||
Db      1 CACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGTTTCCTG 51


RESULT 11
US-08-522-421-1
; Sequence 1, Application US/08522421
; Patent No. 5908973
;  GENERAL INFORMATION:
;    APPLICANT:  Abu-Bakar, Umi Kalsom
;    APPLICANT:  Barton, Sarah Louise
;    APPLICANT:  Gallego-Veigas, Pedro Pablo
;    APPLICANT:  Gray, Julie Elizabeth
;    APPLICANT:  Grierson, Donald
;    APPLICANT:  Lowe, Alexandra Louise
;    APPLICANT:  Picton, Steve
;    APPLICANT:  Whotton, Lee Colin
;    TITLE OF INVENTION:  DNA, DNA CONSTRUCTS, CELLS AND PLANTS
;    TITLE OF INVENTION:  DERIVED THEREFROM
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/522,421
;      FILING DATE:  11-JAN-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305868.3
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305869.1
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305859.2
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305865.9
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305866.7
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305867.5
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305860.0
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305862.6
;      FILING DATE:  22-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9314351.9
;      FILING DATE:  12-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9320988.0
;      FILING DATE:  12-OCT-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  ERT1B
US-08-522-421-1


  Query Match             2.5%;  Score 42.2;  DB 2;  Length 1669;
  Best Local Similarity   49.1%;  Pred. No. 0.019;
  Matches  141;  Conservative    0;  Mismatches  143;  Indels    3;  Gaps    1;

Qy   1081 aagtggatacccagaatgaccttctaggtcatccaaaaaccagagctttataactcat 1140
          || |||  ||| || ||  |  | || |  |||||          ||  ||| | ||||||
Db    972 AAATGGGCACCACAAAAACAGGTGCTTGCACATCCCGCGGTAGCAGGGTTTTTCACTCAT 1031

Qy   1141 ggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattcca 1200
           ||||    |||   |    |||    || |   | |  ||||||||||| |||  |||
Db   1032 TGTGGTTGGAATTCTACGCTTGAAAGTATATGTGAAGAAGTCCCTATGGTGTGCAGGCCA 1091
```

```
Qy   1201 ttgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttaga 1260
          ||  |    || ||||  |  |||       |  |||    | |   |      ||| |
Db   1092 TTTCTAGCAGACCAACTGGTGAACGCAAGGTATCTGAG---CCAAATATACAAGGTTGGG 1148

Qy   1261 ttggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaat 1320
          || || ||  |    ||  ||| || | | |  |||  || | |  | | | ||
Db   1149 TTCGAATTGGAGGTTATCGAGAGAACGGTCATAGAGAAAACAATAAGAAAACTCATGTTA 1208

Qy   1321 gatcctttatataaagagaatattatgaaattatcaagaattcaaca 1367
            |    |  |||||   |  |  | | || ||  ||  ||||
Db   1209 AGTGAAGAAGGCAAAGATGTGAAGAAAAGAGTAGCAGACATGAAACA 1255


RESULT 12
5180581-1
;Patent No. 5180581
;    APPLICANT: MILLER, LOIS K.;O'REILLY, DAVID R.
;    TITLE OF INVENTION: BIOLOGICAL INSECT CONTROL AGENTS AND
;METHODS OF USE
;    NUMBER OF SEQUENCES: 2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/373,952
;      FILING DATE: 29-JUN-1989
;SEQ ID NO:1:
;      LENGTH: 1800
5180581-1


  Query Match             2.5%;  Score 42.2;  DB 6;  Length 1800;
  Best Local Similarity   48.9%;  Pred. No. 0.02;
  Matches  113;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy   1077 gtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1136
          | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
Db   1186 gcaaaattggtttaatcaacgcgccgtgctgcgtcataaaaaaatggcggcgtttattac 1245

Qy   1137 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1196
          || || |||    |    | | | |||||  |  |    |||||  || ||||||  |  |
Db   1246 gcaaggcggactacaatcgagcgacgaggccttggaagccgggataccc atggtgtgtct 1305

Qy   1197 tccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1256
          ||  || |  | || ||   | |  |       |  |  ||      || | |||
Db   1306 gcccatgatgggcgaccagttttaccatgcgcacaaattacagcaactcggcgtagcccg 1365

Qy   1257 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1307
            ||||||      ||  | || ||     | || ||    ||  | ||
Db   1366 cgccttggacactgttaccgtttccagcgatcaactactagtggcgataaa 1416


RESULT 13
US-08-281-916-5
; Sequence 5, Application US/08281916
; Patent No. 5662897
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Lois K.
;    APPLICANT:  Black, Bruce C.
;    APPLICANT:  Dierks, Peter M.
;    APPLICANT:  Fleming, Nancy C.
;    APPLICANT:  Ahmed, Fakhruddin
;    TITLE OF INVENTION:  Insect Viruses, Sequences, Insecticidal
;    TITLE OF INVENTION:  Compositions and Methods of Use
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/281,916
;      FILING DATE:  27-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  29-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
US-08-281-916-5


  Query Match             2.5%;  Score 42.2;  DB 1;  Length 2793;
  Best Local Similarity   48.9%;  Pred. No. 0.025;
  Matches  113;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy   1077 gtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1136
          | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
Db   1186 GCAAAATTGGTTTAATCAACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTATTAC 1245

Qy   1137 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1196
          || || |||    |    | | | |||||  |  |    |||||  || ||||||  |  |
Db   1246 GCAAGGCGGACTACAATCGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCT 1305

Qy   1197 tccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1256
          ||  || |  | || ||   | |  |       |  |  ||      || | |||
Db   1306 GCCCATGATGGGCGACCAGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCG 1365

Qy   1257 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1307
            ||||||      ||  | || ||     | || ||    ||  | ||
Db   1366 CGCCTTGGACACTGTTACCGTTTCCAGCGATCAACTACTAGTGGCGATAAA 1416


RESULT 14
US-08-460-725-7
; Sequence 7, Application US/08460725
; Patent No. 5858353
;  GENERAL INFORMATION:
;    APPLICANT:  Miller, Lois K.
;    APPLICANT:  Black, Bruce C.
;    APPLICANT:  Dierks, Peter M.
;    APPLICANT:  Fleming, Nancy C.
;    APPLICANT:  Ahmed, Fakhruddin
;    TITLE OF INVENTION:  Insect Viruses, Sequences, Insecticidal
;    TITLE OF INVENTION:  Compositions and Methods
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,725
```

```
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/281,916
;      FILING DATE:  27-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  29-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
US-08-460-725-7


  Query Match             2.5%;  Score 42.2;  DB 2;  Length 2793;
  Best Local Similarity   48.9%;  Pred. No. 0.025;
  Matches  113;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy    1077 gtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagctttttataac 1136
           | | || ||| |    ||    | | | ||  |||||  |||||     || ||||| ||
Db    1186 GCAAAATTGGTTTAATCAACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTTATTAC 1245

Qy    1137 tcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcat 1196
           || || |||    |    | | | |||||  |  |    |||||  || ||||||  |  |
Db    1246 GCAAGGCGGACTACAATCGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCT 1305

Qy    1197 tccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgt 1256
           ||  || |  | || ||   | |  |        |  |  ||      || | |||
Db    1306 GCCCATGATGGGCGACCAGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCG 1365

Qy    1257 tagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaa 1307
             ||||||      ||  | || ||     | || ||    ||  | ||
Db    1366 CGCCTTGGACACTGTTACCGTTTCCAGCGATCAACTACTAGTGGCGATAAA 1416


RESULT  15
PCT-US92-00282-18
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18


  Query Match             2.3%;  Score 39.6;  DB 5;  Length 1190;
  Best Local Similarity   47.5%;  Pred. No. 0.089;
  Matches  151;  Conservative    0;  Mismatches  164;  Indels    3;  Gaps    1;

Qy     584 ccctccttcctacatacctattgttatgtcaaaattaagtgatcaaatgactttcatgga 643
           |||    |||||| | || |    | | ||          ||||| ||||| ||| || |
Db     648 CCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACATGACCTTCCTGCA 707

Qy     644 gagggtaaaaaatatgatctatgtgctttattttgacttttggttccaaatgtctgatat 703
           | ||||| || || ||| ||  ||   |||      |||||  || | |  ||  | ||
Db     708 GCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTTTCTGTGCGACGTGGTTTATTC 767

Qy     704 gaagaagtgggatcagttttacagtgaagttttaggaagacccactaccttatttgagac 763
           | |       |   ||  |   ||| |  |    |||      ||  |    ||
Db     768 CCCGTATGCAACCC---TTGCCTCAGAATTCCTTCAGAGAGAGGTGACTGTCCAGGACCT 824

Qy     764 aatgggaaaagctgacatatggcttatgcgaaactcctggagttttcaatttcctcatcc 823
           | || |    ||     | |||||  |  |||    ||          |    |||   ||
Db     825 ATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGATGACCCTAGGCC 884

Qy     824 attcttaccaaacgttgattttgttggaggattccactggcaaacctgccaaacccctac 883
           || | || ||  | | |||||||||| ||| || |||| |    |    |  | | |
Db     885 CATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCAAAATCCACTATC 944

Qy     884 ctaaggaaatggaggagt 901
           | | |    |   |||||
Db     945 CCAGGTGTGTATTGGAGT 962


Search completed: May  1, 2001, 00:31:27
Job time: 56431 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:17:15 ; Search time 12176.5 Seconds
(without alignments)
20.562 Million cell updates/sec

Title: US-09-721-183-2
Perfect score: 1713
Sequence: 1 atcgcattgcaccaggatga.....gatttctttcttcctgtgac 1713

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	690.8	40.3	823	151	BF689099	BF689099 602185172
	2	572.2	33.4	746	151	BF688309	BF688309 602185172
	3	533.6	31.2	1005	174	BG167171	BG167171 602344696
	4	445.2	26.0	513	144	BF105272	BF105272 601822259
c	5	439.6	25.7	761	16	AI118428	AI118428 ue40e09.x
	6	414.8	24.2	651	4	AA255327	AA255327 mz84f10.r
	7	412.2	24.1	923	21	AI529976	AI529976 ui87g04.y
	8	412	24.1	1192	169	BF789664	BF789664 602105279
c	9	408.2	23.8	669	102	AI821970	AI821970 ob22c12.x
	10	396.2	23.1	996	21	AI530408	AI530408 ui92f08.y
	11	394	23.0	616	4	AA255225	AA255225 mz85d04.r
c	12	392.4	22.9	482	14	AA987868	AA987868 oq81f04.s
c	13	386.6	22.6	517	117	AW604423	AW604423 RC3-CT025
	14	385	22.5	792	169	BF786091	BF786091 602110662
	15	382.6	22.3	554	161	BE013295	BE013295 123199 MA
	16	380	22.2	410	31	AV696157	AV696157 AV696157
	17	377.8	22.1	628	31	AV657530	AV657530 AV657530
	18	377.6	22.0	975	149	BF533975	BF533975 602075137

	19	375.2	21.9	934	169	BF786200	BF786200 602110490
	20	363.6	21.2	924	21	AI528431	AI528431 ui97d02.y
	21	363.2	21.2	819	16	AI156939	AI156939 ui49c01.y
c	22	362.2	21.1	432	14	AA962194	AA962194 oq69h06.s
	23	359	21.0	909	21	AI528459	AI528459 ui97f08.y
	24	358.4	20.9	762	11	AA717701	AA717701 vp94g02.r
c	25	358.2	20.9	639	146	BF322562	BF322562 maa28e12.
c	26	357.8	20.9	611	204	AQ315601	AQ315601 RPCI11-97
	27	357.2	20.9	781	21	AI529761	AI529761 ui82b06.y
	28	348.6	20.4	796	16	AI097842	AI097842 ue40e09.y
c	29	338.8	19.8	440	111	AW129268	AW129268 xf22g02.x
c	30	333	19.4	425	165	BE326864	BE326864 hr66b05.x
	31	331.4	19.3	557	143	BF080461	BF080461 231360 MA
c	32	331	19.3	601	205	AQ376853	AQ376853 RPCI-11-1
c	33	331	19.3	643	102	AI785430	AI785430 uj42b06.x
c	34	329.6	19.2	836	21	AI529553	AI529553 ui82a02.x
	35	328.6	19.2	777	16	AI116747	AI116747 ue28e05.y
	36	325.4	19.0	1096	169	BF788115	BF788115 602113339
c	37	324.2	18.9	555	18	AI265736	AI265736 uj05f12.x
c	38	324	18.9	565	11	AA746229	AA746229 ob22c12.s
	39	321.6	18.8	483	14	AF202345	AF202345 AF202345
	40	321.4	18.8	572	121	AW917402	AW917402 EST348706
c	41	321.2	18.8	598	18	AI303476	AI303476 ui71e01.x
c	42	319.6	18.7	589	18	AI256545	AI256545 ui87g04.x
	43	319.4	18.6	453	32	AV719818	AV719818 AV719818
	44	318.8	18.6	966	143	BF027320	BF027320 601671692
	45	311.6	18.2	844	22	AI573962	AI573962 uj66f05.y

ALIGNMENTS

RESULT 1
BF689099/c
LOCUS BF689099 823 bp mRNA EST 22-DEC-2000
DEFINITION 602185172T1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299880 3', mRNA sequence.
ACCESSION BF689099
VERSION BF689099.1 GI:11974507
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 823)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1157 row: o column: 17
High quality sequence start: 5
High quality sequence stop: 796.
FEATURES Location/Qualifiers
source 1. .823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4299880"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. |"
BASE COUNT 221 a 199 c 164 g 239 t
ORIGIN

Query Match 40.3%; Score 690.8; DB 151; Length 823;
Best Local Similarity 91.6%; Pred. No. 1.9e-170;
Matches 753; Conservative 0; Mismatches 67; Indels 2; Gaps 2;

```
Qy    845 tgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggagtttg 904
          | |||  |||  |  |  |||||||  |||||  |||   |  |  |||||||||  ||  ||||
Db    822 TTTTGTTGGAGACTCCACGCAAACCGGCCAACCCCTGCCTAAGGGAAATGGAAGACTTTG 763

Qy    905 tacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaaca 964
          |||||||  |||||||||||||||||||||||||||||||||||||||||||  | |  |||||||
Db    762 TACAGAGTTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACA 703

Qy    965 tgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttc 1024
          ||||||  |||||||||||  ||||||||  |||  |||  |||  ||||||||||||||||||
Db    702 TGACAGAAGAAAGGGCCAACGTAATTGCATCAGACCTGGCCCAGATCCCACAAAAGGTTC 643

Qy   1025 tgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataagt 1084
          ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    642 TGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGT 583

Qy   1085 ggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatggtg 1144
          ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db    582 GGATACCCCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTG 523

Qy   1145 gagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccattgt 1204
          ||||||||||||||||  ||||||||||||||||||||||||||||||||  ||||||||||
Db    522 GAGCCAATGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGT 463

Qy   1205 tttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagattgg 1264
          ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db    462 TTGCCGATCAACCTGATAACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGG 403

Qy   1265 acttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatc 1324
          ||||||||||||||||||||||||||  ||||||||||  ||||||  ||||||||||||||
Db    402 ACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATC 343

Qy   1325 ctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagc 1384
          |||  ||||||||||||| ||||||||||||||||||||||||||||||||||||||  |||
Db    342 CTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAG- 282

Qy   1385 ccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacacc 1444
          ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||| |||||||
Db    283 CCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACC 224

Qy   1445 ttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgggt 1504
          ||||  ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    223 TTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGT 164

Qy   1505 ttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttgtt 1564
          | |||||||  |||||||||||||||||||||||||||  |||||||  | ||||||||||||
Db    163 TCCTGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTT 104

Qy   1565 tctggaagtttgctagaaaagggaagaagggaaaaagagattagttatgtctgacatttg 1624
          ||||||||||||||||||  ||||||||||||  |||||  ||||||||||||||||||||||
Db    103 TCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTATATCTGAGATTTG 44

Qy   1625 aagctggaaaaccaga-tagataggacaacttcagtttattc 1665
          ||||||||||||||  ||  |||  |  |||  |||||||||  |  |
Db     43 AAGCTGGAAAACCTGACTAGGTGAGACTACTTCAGTTATTCC 2
```

RESULT 2
BF688309
LOCUS BF688309 746 bp mRNA EST 22-DEC-2000
DEFINITION 602185172F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299880 5',

```
            mRNA sequence.
ACCESSION   BF688309
VERSION     BF688309.1  GI:11973717
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement:.ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1157  row: o  column: 17
            High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4299880"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      212 a    158 c    183 g    193 t
ORIGIN


  Query Match             33.4%;  Score 572.2;  DB 151;  Length 746;
  Best Local Similarity   92.5%;  Pred. No. 2.5e-139;
  Matches  667;  Conservative    0;  Mismatches   48;  Indels    6;  Gaps    6;

Qy    843 tttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggagtt 902
          ||||||||||||| |||||| |||||||||||||||||||| |||||||||||||| || ||
Db      3 TTTGTTGGAGGACTCCACT-GCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTT 61

Qy    903 tgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaa 962
          |||||||||||||||||||||||||||||||||||||||||||||||||| || | |||||
Db     62 TGTACAGAGCTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAA 121

Qy    963 catgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggt 1022
          |||||||| |||||||||| ||||||||| ||||||| ||| ||||||||||||||||
Db    122 CATGACAGAAGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGT 181

Qy   1023 tctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgtataa 1082
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    182 TCTGTGGAGATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAA 241

Qy   1083 gtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactcatgg 1142
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    242 GTGGATACCCCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGG 301

Qy   1143 tggagccaatggcatctatgaggcaatctaccatgggatccctatggtgggcattccatt 1202
          |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db    302 TGGAGCCAATGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATT 361

Qy   1203 gttttgggatcaacctgataacattgctcacatgaaggccaagggagcagctgttagatt 1262
          ||||   |||||||||||||||||||||||||||||||| |||||||||||||||| |
Db    362 GTTTGCCGATCAACCTGATAACATTGCTCACATGAAGGCC-AGGGAGCAGCTGTTAGAGT 420

Qy   1263 ggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagtaattaatga 1322
          |||||||||||||||||||||||||||||||||| |||||||||| |||||| ||||||||||
Db    421 GGACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGA 480

Qy   1323 tcctttatataaagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaa 1382
          ||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||| ||
Db    481 TCCTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAA 540

Qy   1383 gcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaaggagccaaaca 1442
          |||||||||||||||||||||||||||| |||||||||||| |||||||||||| |||||
Db    541 GCCCCTGGATCGAGCAGTCTTCTGGATCGAATTTGTCATGCGCCACAAAGGAGCTAAACA 600

Qy   1443 ccttcgagttgcagcccatgacctcacctggttccagtaccactctttggatgtgattgg 1502
          |||||| |||||||||||| || |||||||||||||||||||||| ||||||||||||||
Db    601 CCTTCGGGTTGCAGCCCACGA-CTCACCTGGTTCCAGTACCACTC-TTGGATGTGATTGG 658

Qy   1503 gtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttttgtctgttttg 1562
          | ||||||||| ||||||||| |||||||||||          || ||| | | | | ||
Db    659 G-TTCTGCTGGTCTGTGTGGC-ACTGTGATATTATCGTCACAAATGTTGTCTGTCTGTTC 716

Qy   1563 t 1563
          |
Db    717 T 717


RESULT   3
BG167171
LOCUS       BG167171     1005 bp    mRNA            EST       06-FEB-2001
DEFINITION  602344696F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4455022 5',
            mRNA sequence.
ACCESSION   BG167171
VERSION     BG167171.1  GI:12673874
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247  row: o  column: 23
            High quality sequence stop: 696.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4455022"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      256 a    252 c    252 g    244 t      1 others
ORIGIN
```

```
  Query Match            31.2%;  Score 533.6;  DB 174;  Length 1005;
  Best Local Similarity  85.2%;  Pred. No. 3.7e-129;
  Matches  655;  Conservative    0;  Mismatches  105;  Indels    9;  Gaps    5;

Qy     735 ttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcga 794
           || ||||||||||||||| |||| |||||||||||| ||||||||| |||||||||| |||
Db     103 TTTGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTATTCGA 162

Qy     795 aactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggagga 854
           |||||||||| |||||| |||||||||||| |||||||||| ||||||||||||||||||
Db     163 AACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATTTTGTTGGAGGA 222

Qy     855 ttccactggcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctc 914
            |||||| |||||||||||||||||| |||||||||||||| || ||||||||||||||
Db     223 CTCCACT-GCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAGCTC 281

Qy     915 tggagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcaga 974
           |||||||||||||||||||||||||||||||||||||| || | |||||||||||||| |||
Db     282 TGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGA 341

Qy     975 aagggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatt 1034
           ||||||||| ||||||||| ||||||| ||| ||||||||||||||||||||||||||||
Db     342 AAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATT 401

Qy    1035 tgatgggaataaaccagatgccttaggtctcaatactcggctgtataagtggatacccca 1094
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     402 TGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACCCCA 461

Qy    1095 gaatgaccttctaggtcatccaaaaaccagagctttttataactcatggtggagccaatgg 1154
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     462 GAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGG 521

Qy    1155 catctatgaggcaatctaccatgggatccctatggtgggcattccattgttttgggatca 1214
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     522 CATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCA 581

Qy    1215 acctgataacattgctcacatgaaggccaagggagcagctgttagattggacttc-aaca 1273
           |||||||||||||||||||||||||||| | |||||||||||||| |||||||| ||||
Db     582 NACTGATAACATTGCTCACATGAAGGCCAGGCGAGCAGCTGTTAGAGTGGACTTCAAACA 641

Qy    1274 caatgtcgagtacagacctgctgaatgcactgaagacagtaattaatgatccttttatata 1333
           |||||||||||||||| |||||||||| ||| || ||||||| ||||||||| ||||
Db     642 CAATGTCGAGTACAGACTTGCTGAATGC-TTGACGAGAGTAATTCATGATCCTTCATATC 700

Qy    1334 aagagaatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatc 1393
           | |||||| |    | | | ||   | |      | ||||         | ||| |||||
Db     701 ACGAGAATGTCTGCACACTCTCCGAATTCACTGGTCATCA-----GTGAAGCCCCGGATC 755

Qy    1394 gagcagtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttg 1453
           |||||| |||| |||||||| ||| |||||  |  || || |  ||||   | |   |||
Db     756 GAGCAGGCTTCCGGATTGAA-TTGGCATGCGGCCAAAGGGCGTAAAACCTTTGCTGCTTG 814

Qy    1454 cagcccatgacctcacctggttccagtaccactctttggatgtgattgg 1502
           |  ||||  | |    | |||||| || || ||||  |    |||| ||
Db     815 CGTCCCAGAAACCTCACCGGTTCCCGTCCCCCTCTGGGAGGTTGATGGG 863


RESULT  4
BF105272
LOCUS       BF105272      513 bp    mRNA            EST       19-OCT-2000
DEFINITION  601822259F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4042383 5',
            mRNA sequence.
ACCESSION   BF105272
VERSION     BF105272.1  GI:10887798
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM863  row: f  column: 16
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4042383"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      147 a    116 c    121 g    129 t
ORIGIN


  Query Match            26.0%;  Score 445.2;  DB 144;  Length 513;
  Best Local Similarity  93.3%;  Pred. No. 4.7e-106;
  Matches  476;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy     737 aggaagacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaa 796
           |||||||||||||||| |||| |||||||||||| ||||||||| |||||||||| |||||
Db       5 AGGAAGACCCACTACATTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTATTCGAAA 64

Qy     797 ctcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggatt 856
           |||||||| |||||| ||||| ||||| |||||||||| |||||||||||||||||||| |
Db      65 CTCCTGGAATTTTCAGTTTCCATATCCACTCTTACCAAATGTTGATTTTGTTGGAGGACT 124

Qy     857 ccactggcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctctg 916
           ||||| |||||||||||||||||| |||||||||||||| || ||||||||||||||||
Db     125 CCACT-GCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAGCTCTG 183

Qy     917 gagaaaatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaa 976
           |||||||||||||||||||||||||||||||||||||| || | |||||||||||| |||||
Db     184 GAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAAA 243

Qy     977 gggccaatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttg 1036
           ||||||| ||||||||| ||||||| ||| |||||||||||||||||||||||||||||
Db     244 GGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTG 303

Qy    1037 atgggaataaaccagatgccttaggtctcaatactcggctgtataagtggataccccaga 1096
           |||||||||||||||||| |||||||||||||||||||||||| || ||||||||||||||
Db     304 ATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTCTACAAGTGGATACCCCAGA 363

Qy    1097 atgaccttctaggtcatccaaaaaccagagctttttataactcatggtggagccaatggca 1156
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     364 ATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGCA 423

Qy    1157 tctatgaggcaatctaccatgggatccctatggtgggcattccattgttttgggatcaac 1216
           |||| |||||||||||||||||||||||||||||| ||||||||||||   |||||||
Db     424 TCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCAAC 483

Qy    1217 ctgataacattgctcacatgaaggccaagg 1246
           |||||||||||||||||||||||||||| ||
Db     484 CTGATAACATTGCTCACATGAAGGCCAGGG 513
```

```
RESULT   5
AI118428/c
LOCUS       AI118428      761 bp    mRNA            EST       02-SEP-1998
DEFINITION  ue40e09.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1482856 3' similar to gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AI118428
VERSION     AI118428.1  GI:3518752
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:931212
            Seq primer: custom primer used
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .761
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1482856"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      198 a    168 c    165 g    230 t
ORIGIN
```

Query Match 25.7%; Score 439.6; DB 16; Length 761;
Best Local Similarity 74.8%; Pred. No. 1.6e-104;
Matches 564; Conservative 0; Mismatches 189; Indels 1; Gaps 1;

```
Qy    874 aaaccccctacctaagga-aatggaggagtttgtacagagctctggagaaaatggtgttgt 932
          |||||| | ||||||||  ||||| |  ||||| |||||||||| |||  ||||||| ||
Db    761 AAACCCTTGCCTAAGGATTATGGAAGGATTTGTCCAGAGCTCTGAAGACCATGGTGTGGT 702

Qy    933 ggtgttttctctggggtcagtgataagtaacatgacagcagaaagggccaatgtaattgc 992
          ||| |||||||| |||||| || | |||||||||||||| |||||  ||||| | ||||||
Db    701 GGTCTTTTCTCTTGGGTCAATGGTTAGTAACATGACAGAAGAAAAAGCCAACGCAATTGC 642

Qy    993 aacagcccttgccaagatcccacaaaaggttctgtggagatttgatgggaataaaccaga 1052
          |   ||||||||| |||| |||||||||||||| |||| ||||||||| || |  ||||
Db    641 ATGGGCCCTTGCCCAGATTCCACAAAAGGTTCTTTGGAAATTTGATGGCAAAACCCCAGC 582

Qy   1053 tgccttaggtctcaatactcggctgtataagtggataccccagaatgaccttctaggtca 1112
            | ||||| | ||||||  |  |||| |||||| | ||||||||||||| || |||||
Db    581 AACTTTAGGACACAATACCAGAGTGTACAAGTGGCTCCCCCAGAATGACCTCCTTGGTCA 522

Qy   1113 tccaaaaaccagagcttttataactcatggtggagccaatggcatctatgaggcaatcta 1172
          |||||||||| ||| ||| |||||||||||| ||||||||  |||| ||||| |||||
Db    521 TCCAAAAACCAAAGCCTTTGTAACTCATGGTGGGGCCAATGGTGTCTACGAGGCGATCTA 462

Qy   1173 ccatgggatccctatggtgggcattccattgttttgggatcaacctgataacattgctca 1232
            ||||| |||||||| | |||||||| |||||| | || || | ||||||||||| ||
Db    461 TCATGGAATCCCTATGATTGGCATTCCTTTGTTTGGAGACCAGCATGATAACATTGCCCA 402

Qy   1233 catgaaggccaagggagcagctgttagattggacttcaacacaatgtcgagtacagacct 1292
            |||  |||||| |||||||||||  |||| |  |||  |||||||| |   |||| |
Db    401 TATGGTGGCCAAAGGAGCAGCTGTTGCATTGAATATCAGAACAATGTCAAAGTCAGATGT 342

Qy   1293 gctgaatgcactgaagacagtaattaatgatcctttatataaagagaatattatgaaatt 1352
            ||| |||||||| ||   || ||  |  ||||||| |||||| |||||  ||||   ||
Db    341 GCTCAATGCACTGGAGGAGGTCATAGAAAATCCTTTCTATAAAAAGAATGCTATGTGGTT 282

Qy   1353 atcaagaattcaacatgatcaaccagtaaagcccctggatcgagcagtcttctggattga 1412
            ||||  ||||| ||||| || ||  | || ||||||||  | || ||||||||||||||
Db    281 GTCAACCATTCACCATGACCAGCCTATGAAACCCCTGGACAGGGCTGTCTTCTGGATTGA 222

Qy   1413 atttgtcatgccccacaaaggagccaaacaccttcgagttgcagcccatgacctcacctg 1472
            ||||| |||| ||||||| ||||||| |||||  ||      |  ||| |||| |||||
Db    221 GTTTGTTATGCGCCACAAAAGAGCCAAGCACCTGAGACCACTTGGACATAACCTTACCTG 162

Qy   1473 gttccagtaccactctttggatgtgattgggtttctgctggcctgtgtggcaactgtgat 1532
            || |||||||||||| |||||||||||| || || ||  | ||||||||||    ||
Db    161 GTACCAGTACCACTCTCTGGATGTGATTGGATTCCTACTCTCTTGTGTGGCAACCACTAT 102

Qy   1533 atttatcatcacaaagttttgtctgttttgtttctggaagtttgctagaaaagggaagaa 1592
            | |  | |    |||||      ||||  || | |   ||||  |  ||||  || ||
Db    101 AGTCCTTAGTGTAAAGTGCCTCTTGTTCATTTACCGATTCTTTGTAAAGAAAGAAAATAA 42

Qy   1593 gggaaaaagagattagttatgtctgacatttgaa 1626
               || |  || |||   |  |||||| |  ||
Db     41 AATGAAGAATGAGTAGAGCTCACTGACAATGAAA 8
```

```
RESULT   6
AA255327
LOCUS       AA255327      651 bp    mRNA            EST       14-MAR-1997
DEFINITION  mz84f10.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720139 5'
            similar to gb:U06641 UDP-GLUCURONOSYLTRANSFERASE 2B8 PRECURSOR,
            MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AA255327
VERSION     AA255327.1  GI:1889924
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
```

```
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:445635
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 416.
FEATURES              Location/Qualifiers
     source           1. .651
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:720139"
                      /clone_lib="Soares mouse NML"
                      /tissue_type="Liver"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT      192 a    144 c    147 g    168 t
ORIGIN


  Query Match             24.2%;  Score 414.8;  DB 4;  Length 651;
  Best Local Similarity   79.4%;  Pred. No. 4.9e-98;
  Matches  516;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy    742 gacccactaccttatttgagacaatgggaaaagctgacatatggcttatgcgaaactcct 801
          | || || |||||| |||||||||||||| ||||| || || ||||| ||  ||  ||  |
Db      1 GGCCAACCACCTTAATTGAGACAATGGGCAAAGCAGAAATGTGGCTCATTAGATCCTATT 60

Qy    802 ggagttttcaatttcctcatccattcttaccaaacgttgattttgttggaggattccact 861
          ||  |||  | |||||||| |||  ||||||||| ||| | | |||||||||| |||| |
Db     61 GGGATTTGGAGTTTCCTCACCCAACCTTACCAAATGTTTACTATGTTGGAGGAGTCCA-T 119

Qy    862 ggcaaacctgccaaacccctacctaaggaaatggaggagtttgtacagagctctggagaa 921
          |||||||||| |||||  | |||||||||||||| || ||||| ||||||||||||||
Db    120 TGCAAACCTGCTAAACCATTGCCTAAGGAAATGGAAGACTTTGTTCAGAGCTCTGGAGAG 179

Qy    922 aatggtgttgtggtgttttctctggggtcagtgataagtaacatgacagcagaaagggcc 981
          ||||||| |||||||||||||||||||||| || | ||||| ||||||| ||||| ||||
Db    180 CATGGTGTGGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAATATGACAGAAGAAAAGGCC 239

Qy    982 aatgtaattgcaacagcccttgccaagatcccacaaaaggttctgtggagatttgatggg 1041
          ||||  |||||||   || ||||||| || |||||||||||||| ||||||||||| ||
Db    240 AATGCAATTGCATGGGCTCTTGCCAGCATTCCACAAAAGGTTCTTTGGAGATTTGACGGC 299

Qy   1042 aataaaccagatgccttaggtctcaatactcggctgtataagtggataccccagaatgac 1101
          || |  ||||   ||||||| | |||||  |  | || |||||| | ||||| |||||
Db    300 AAAACTCCAGCCACCTTAGGACCCAATACCAGAATCTACAAGTGGCTTCCCCAAAATGAT 359

Qy   1102 cttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatctat 1161
          ||||| |||||||||||||||| ||| |||||||||||||||| || |||||  ||||
Db    360 CTTCTTGGTCATCCAAAAACCAAAGCCTTTATAACTCATGGTGGCGCGAATGGACTCTAC 419

Qy   1162 gaggcaatctaccatgggatccctatggtgggcattccattgttttgggatcaacctgat 1221
          ||||| ||| | ||||| |||||||| | |||||||| |||||| | || |||| ||||
Db    420 GAGGCGATCCATCATGGAATCCCTATGATTGGCATTCCTTTGTTTGGAGAACAACATGAT 479

Qy   1222 aacattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatgtcg 1281
          ||||||||||||||  |||||| |||||||||||| |||| |  |||  || ||||||
Db    480 AACATTGCTCACATGGTGGCCAAAGGAGCAGCTGTTACATTGAATATCAGAACGATGTCA 539

Qy   1282 agtacagacctgctgaatgcactgaagacagtaattaatgatccttttatataaagagaat 1341
          ||  ||||  |||| ||||||||| |   ||| ||  |  ||||||| ||||||||||||
Db    540 AGGTCAGATTTGCTCAATGCACTGGACGAAGTCATAGACAATCCTTTCTATAAAGAGAAT 599

Qy   1342 attatgaaattatcaagaattcaacatgatcaaccagtaaagcccctgga 1391
            ||||   || || |  |||||||||| || ||  | ||||||||||
```

```
Db    600 GCTATGTGGTTGTC-ACCATTCAACATGACCAGCCTATGAAGCCCCTGGA 648


RESULT   7
AI529976
LOCUS       AI529976      923 bp    mRNA            EST       18-MAR-1999
DEFINITION  ui87g04.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1889430 5' similar to gb:Y00317 UDP-GLUCURONOSYLTRANSFERASE
            2B4 PRECURSOR, MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AI529976
VERSION     AI529976.1  GI:4444111
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:973754
            Seq primer: custom primer used
            High quality sequence stop: 492.
FEATURES              Location/Qualifiers
     source           1. .923
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1889430"
                      /clone_lib="Sugano mouse liver mlia"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                      (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                      was primed with an oligo(dT) primer
                      [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                      ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                      and cloned into distinct DraIII sites of the pME18S-FL3
                      vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                      be used to isolate the cDNA insert.  Size selection was
                      performed to exclude fragments <1.5kb.  Library
                      constructed by Dr. Sumio Sugano (University of Tokyo
                      Institute of Medical Science).  Custom primers  for
                      sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                      primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      273 a    181 c    189 g    277 t      3 others
ORIGIN


  Query Match             24.1%;  Score 412.2;  DB 21;  Length 923;
  Best Local Similarity   69.9%;  Pred. No. 2.6e-97;
  Matches  597;  Conservative    0;  Mismatches  251;  Indels    6;  Gaps    3;

Qy      7 ttgcaccaggatgactctgaaatggacttcagttcttctgctgatacatctccagttgtt 66
          || || || |||| || |||||  | |||||||| ||||| |||||||  | ||  || |
Db     21 TTTCAGCAAGATGTCTATGAAACAGGCTTCAGTTTTTCTGTTGATACA-GTTCATATGCT 79

Qy     67 actttagctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattgga 126
          |  ||||  ||||    ||||| |||||||||||||| |  ||||||||||||||||||
```

```
Db      80 ATATTAGACCTGGAGCCTGTGGGAAAGTGCTGGTGTGGCCTACAGAATACAGCCATTGGA 139

Qy     127 tgaatatgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactgg 186
           | |||||||| | ||||||| | || ||||| ||||||||||||| || || || ||
Db     140 TAAATATGAAAATAATCCTGGATGAACTTGTCCAGAGAGGTCATGACGTCACCGTTCTCA 199

Qy     187 catcttcagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagttt 246
           |||||| |||||||| ||  ||| || |  ||| |||  || |||| |||||| |||
Db     200 TATCTTCTGCTTCCATCCTCATTGGGCCGAGCAATGAATCTTCTATTAATTTTGAAATTT 259

Qy     247 atcctacatctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagat 306
           || || || |||| | ||||  |||  ||||  ||    |     ||  |||    | |
Db     260 ATTCTGCACCTTTGAGTAAAGATGATCTTGAATATGCTTTTGAAAAATGGGTAGGAAACT 319

Qy     307 ggtca---gacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcc 363
           || ||   ||  |   |||| |  |||||| || || |||| ||   ||| |||||
Db     320 GGACATACGAATTAAAAAAACTTCCATTTTGGACATCTTATTCAAAACTGCAAAAAATCT 379

Qy     364 tgtgggaattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaag 423
           | ||||  |  |||||  ||  ||  ||||| ||||  ||||||| || ||||
Db     380 CCAGTGAATATTCAGACATGATTGAAAGTTTCTGCAAAGCAGTAGTTTGGAACAAGAGCC 439

Qy     424 ttatgaaaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccct 483
           | |||||||||||| ||||  || |  |||||  ||||| | ||||||||  |  ||||||
Db     440 TCATGAAAAAACTCCAAGGATCTAAGTTTGATGTCGTTCTAGCAGATGCCTTGGTTCCCT 499

Qy     484 gtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactc 543
           ||||||||||||  | |  || |||||  | ||  || || ||||||||||||||  ||
Db     500 GTGGTGAGCTGCTATCAGAACTGCTTAAGACACCTTTAGTATACAGTCTCCGCTTCTGTC 559

Qy     544 ctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacataccta 603
           |||  |||| |  ||||| | ||||||| || ||     |||| ||||||||  | |||
Db     560 CTGCATACAAATGTGAAAAGTACAGTGGGGGCCTTTCACTCCCCCCTTCCTATGTGCCTG 619

Qy     604 ttgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatct 663
           | ||| ||||| || ||||||| || ||||| ||     ||   ||| || |||||| |
Db     620 TGGTTCTGTCAGAACTAAGTGACCACATGACATTTGCAGAANAGGTGAAGAATATGTTGC 679

Qy     664 atgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggatcagtttt 723
           | ||| |   | |||||||||||| | || |  | | |   |||   ||| |||||||||
Db     680 ANGTGTTGCTTCTTGACTTTTGGCTTCANACATTTAACGAGAAATCCTGGAATCAGTTTT 739

Qy     724 acagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatat 783
           ||||||| ||| |||| ||||| || || |||  ||||  ||||| |||| |||||||
Db     740 ACAGTGATGTTCTAGGGAGACCTACAACATTAACTGAGATGATGGGGAAAGCAGACATAT 799

Qy     784 ggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgatt 843
           ||||    |||| || |||||    | || || |||| || ||||| || ||  |||| |
Db     800 GGCTCGGTCGAACCTTCTGGA--CCTGAAATTTCTCACCCTTTCTTGCCTAATTTTGACT 857

Qy     844 ttgttggaggattc 857
           ||| |||||  |||
Db     858 TTGGTGGAGACTTC 871


RESULT   8
BF789664
LOCUS       BF789664     1192 bp    mRNA            EST       12-JAN-2001
DEFINITION  602105279F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4223443
            5', mRNA sequence.
ACCESSION   BF789664
VERSION     BF789664.1  GI:12094700
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1192)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9811  row: n  column: 20
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1. .1192
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4223443"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      277 a    304 c    276 g    335 t
ORIGIN


  Query Match           24.1%;  Score 412;  DB 169;  Length 1192;
  Best Local Similarity   77.3%;  Pred. No. 3.2e-97;
  Matches  538;  Conservative    0;  Mismatches  155;  Indels    3;  Gaps    3;

Qy     834 aacgttgattttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaat 893
           || ||||| | |||||||||| |||||| |||||||||| |||||  | ||||||| ||
Db       1 AATGTTGACTATGTTGGAGGACTCCACT-GCAAACCTGCTAAACCTTTGCCTAAGGATAT 59

Qy     894 ggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagt 953
           ||| || ||||| |||||||||||||  ||||||| |||||||||| || |||||| |
Db      60 GGAAGAATTTGTCCAGAGCTCTGGAGACCATGGTGTGGTGGTGTTTTCACTTGGGTCAAT 119

Qy     954 gataagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatccc 1013
           | | |||||||||||| ||||| |||||  |||||||  |||||||| |||| ||
Db     120 GGTTAGTAACATGACAGAAGAAAAAGCCAACACAATTGCATGGGCCCTTGCCCAGATTCC 179

Qy    1014 acaaaaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcg 1073
           |||||||||||| |||| ||||||||| || |  ||||   | |||||  | ||||||  |
Db     180 ACAAAAGGTTCTTTGGAAATTTGATGGCAAAACCCCAGCAACTTTAGGACACAATACGAG 239

Qy    1074 gctgtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttat 1133
             |||| |||||| | |||||||||||||| || ||||||||||||| | ||| ||| |
Db     240 AGTGTACAAGTGGCTCCCCCAGAATGACCTCCTTGGTCATCCAAAAACAAAAGCCTTTGT 299

Qy    1134 aactcatggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtggg 1193
           |||||||||||| |||||||| |||||||| |||||  |||| |||||||||  | ||
Db     300 AACTCATGGTGGGGCCAATGGTGTCTATGAGGTGATCTATCATGGAATCCCTATGATTGG 359

Qy    1194 cattccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcagc 1253
           |||||| |||||| | || |||| |||||| ||||| || |||  |||||| ||||||||
Db     360 CATTCCTTTGTTTGGAGAACAACATGATAATATTGCCCATATGGTGGCCAAAGGAGCAGC 419

Qy    1254 tgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacagt 1313
           ||||| |||| |  |||  |||||||| ||  ||||  | || ||||||||| ||  |||
Db     420 TGTTACATTGAATATCAGAACAATGTCAAGATCAGATGTACTCAATGCACTGGAGGAAGT 479

Qy    1314 aattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgatca 1373
           ||  |  |||||||| ||||| |||    || ||||  ||||| || || ||
Db     480 CATAGACAATCCTTTCTATAAAAAGAATGCTATATGGTTGTCAACCATTCACCACGACCA 539

Qy    1374 accagtaaagcccctggatcgagcagtcttctggattgaatttgtca-tgccccacaaag 1432
           ||    ||||||||||  | || |||||||| |||| ||||||| ||| ||||||||
Db     540 GCCTACGAAGCCCCTGGACAGGGCTGTCTTCTGGGTTGAGTTTGTCATTGCGCCACAAAA 599

Qy    1433 gagccaaacaccttcgagttgcagccca-tgacctcacctggttccagtaccactctttg 1491
           | || || |||||  ||      |  || | |||| ||||||  ||||||||| | | |
Db     600 GGGCTAAGCACCTGAGATCACTTGGACATTAACCTTACCTGGCACCAGTACCAATTTCTA 659
```

```
Qy   1492 gatgtgattgggtttctgctggcctgtgtggcaact 1527
          |||||||| || || || ||  | ||||| || | |
Db    660 GATGTGATCGGCTTCCTACTCTCTTGTGTTGCCAGT 695


RESULT  9
AI821970/c
LOCUS       AI821970      669 bp    mRNA            EST       13-DEC-1999
DEFINITION  ob22c12.x5 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1324438 3'
            similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION   AI821970
VERSION     AI821970.1  GI:5441049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ob22c12.y5
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: National Cancer Institute, Cancer Genome
            Anatomy Project (CGAP), Tumor Gene Index
            This read has been verified (found to hit its original self in the
            correct orientation)
            Insert Length: 1566   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1324438"
                     /clone_lib="NCI_CGAP_Kid5"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGCAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      189 a    137 c    108 g    235 t
ORIGIN


  Query Match           23.8%;  Score 408.2;  DB 102;  Length 669;
  Best Local Similarity  90.4%;  Pred. No. 2.6e-96;
  Matches  471;  Conservative    0;  Mismatches   43;  Indels    7;  Gaps    3;

Qy   1193 gcattccattgttttgggatcaacctgataacattgctcacatgaaggccaagggagcag 1252
          | ||| |||||||    ||||||| |||||||| |||||||||| ||| || ||||||||
Db    669 GGATTACATTGTTGACCGATCAACTTGATAACA-TGCTCACATG-AGGGCAGGGGAGCAG 612

Qy   1253 ctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagacag 1312
          |||||||| |||||||||||||| |||||||||||||| |||||||||| |||||| ||
Db    611 CTGTTAGAGTGGACTTCAACACGATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAG 552

Qy   1313 taattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgatc 1372
          ||| |||||||||| |||||||||||||| | ||||||||||||||||||||||||||||
Db    551 TAAATAATGATCCTTCATATAAAGAGAATGTAATGAAATTATCAAGAATTCAACATGATC 492

Qy   1373 aaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaag 1432
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    491 AACCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAG 432

Qy   1433 gagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttgg 1492
          |||| |||||| |||| |||||||||| |||||||||||||||||||||||||||||||
Db    431 GAGCTAAACACATTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGG 372

Qy   1493 atgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagtttt 1552
          |||||||||||||  |||||| ||||||||||||||||||||||| ||||||| | ||
Db    371 ATGTGATTGGGTTCATGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTT 312

Qy   1553 gtctgttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagttat 1612
          |||||||||||||||||||||||||||||||  |||||||||||||  ||||||||||
Db    311 GTCTGTTTTGTTTCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTAT 252

Qy   1613 gtctgacatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaag 1672
          |||||  |||||||||||||||||  |||||  |  |||  ||||||||||||||||||||
Db    251 ATCTGAAATTTGAAGCTGGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAAG 192

Qy   1673 aaagaaaagattgttatgcaagatttctttcttcctgtgac 1713
               ||||||||| |||||||||||||||||||||||| |||
Db    191 -----AAAGATTGTGATGCAAGATTTCTTTCTTCCTGAGAC 156


RESULT  10
AI530408
LOCUS       AI530408      996 bp    mRNA            EST       18-MAR-1999
DEFINITION  ui92f08.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1889895 5' similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE
            2B7 PRECURSOR, MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for
            UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION   AI530408
VERSION     AI530408.1  GI:4444543
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:974219
            Seq primer: custom primer used
            High quality sequence stop: 527.
FEATURES             Location/Qualifiers
     source          1. .996
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1889895"
```

/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

BASE COUNT 273 a 180 c 223 g 316 t 4 others
ORIGIN

Query Match 23.1%; Score 396.2; DB 21; Length 996;
Best Local Similarity 66.3%; Pred. No. 4.2e-93;
Matches 629; Conservative 0; Mismatches 312; Indels 8; Gaps 4;

```
Qy     13 caggatgactctgaaatggacttcagttcttctgctgatacatctccagttgttacttta 72
          || |||| ||   || |||| ||| | ||| |||||| | ||  |  ||||| | ||| |
Db     39 CAAGATGCCTGGAAAGTGGATTTCTGCTCTGCTGCTGCTGCAGAT-AAGTTGCTGCTTCA 97

Qy     73 gctctgggagttgtggaaaagtgctggtgtgggccgcagaatacagccattggatgaata 132
          | |||| ||  ||||| ||||| |||||||| |    |||| ||| ||||||||||||
Db     98 GATCTGTGAAATGTGGGAAAGTGTTGGTGTGGCCGATGGAATTCAGTCATTGGATGAATA 157

Qy    133 tgaagacaatcctgaaagagcttgttcagagaggtcatgaggtgactgtactggcatctt 192
          | || | ||| ||| | || || || |||||||| ||||| || ||||| ||| | |||
Db    158 TAAAAATAATACTGGATGAACTGGTACAGAGAGGCCATGAAGTCACTGTTCTGAGACCTT 217

Qy    193 cagcttccattctttttgatcccaatgatgcatccactcttaaatttgaagtttatccta 252
          |||||| |  | || |||||||||||  |  || |    ||||| ||||||  || |||||
Db    218 CAGCTTACTATGTTCTTGATCCCAAAAAATCACCTGGCCTTAAGTTTGAAACTTTTCCTA 277

Qy    253 catctttaactaaaactgaatttgagaatatcatcatgcaacaggttaagagatgg---t 309
          ||||| | | ||||    |  | || ||| |  ||||  |    ||  |    |||   |
Db    278 CATCTGTCAGTAAAGACAATCTGGAAAATTTTTTCATTAAGTTTGTGGATGTGTGGACTT 337

Qy    310 cagacattcgaaaagatagcttttggttatatttttcacaagaacaagaaatcctgtggg 369
          || || | ||  ||||  | || || ||||| | |   |   ||| | ||  |   |
Db    338 ATGAAATGCCAAGGGATACATGTTTGTCATATTCTCCTTTACTGCAAAATATGATTGATG 397

Qy    370 aattatatgacatatttagaaacttctgtaaagatgtagtttcaaataagaaagttatga 429
          |||| | |||    |||  ||   | ||||||||||| |||||||| ||  |  | ||||
Db    398 AATTTTCTGATTATTTTCTAAGTCTTTGTAAAGATGTTGTTTCAAACAAAGAGCTCATGA 457

Qy    430 aaaaactacaagagtcaagatttgacatcgtttttgcagatgctgtttttccctgtggtg 489
           ||| ||||| || || |  |||||  || ||||  ||||| |||||  |  |||||| |
Db    458 CAAAGCTACAGGAATCCAAGTTTGATGTCCTTTTGTCAGATCCTGTTGCTTTCTGTGGNG 517

Qy    490 agctgctggctgcgctacttaacatacggtttgtgtacagtctccgctttactcctggct 549
          ||||  | ||||  || ||  | || |  ||| |||||||| |||||||  |||| ||||
Db    518 AGCTCATAGCTGAACTGCTCCAGATTCCTTTTCTGTACAGTATCCGCTTCTCTCCAGGCT 577

Qy    550 acacaattgaaaggcacagtggaggactgattttccctccttcctacatacctattgtta 609
          ||  ||||||||    ||| ||  || |  |  ||| ||| || ||  ||||| |  ||
Db    578 ACCAAATTGAAAAATCCAGCGGGAGATTTCTACTCCGTCCCTCTTATGTACCTGTNATTC 637

Qy    610 tgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatatgatctatgtgc 669
          |||||  |||| |||  |||||||| ||||| |||||| ||||||||||||| |   |||
Db    638 TGTCAGGATTAGGTGGACAAATGACATTCATAGAGAGGATAAAAAATATGATATGCATGC 697

Qy    670 tttatttttgacttttggttccaaatgtctgatatgaagaagtgggatcagttttacagtg 729
          | || |||||||| | ||||| |||| | |     |||| |||||    ||||||||||
Db    698 TNTACTTTGACTTTGGCTTCCAGATGTTTAAATGATAGAAATGGGACTCATTTTACAGTG 757
```

```
Qy    730 aagttttaggaagacccactaccttatttgagacaatgggaaaagctgacatatggctta 789
          |   ||| ||||| || || |||||| ||||||| |||||  ||||    || ||||||
Db    758 AGTATTTGGGAAGGCCAACCACCTTAATTGAGACCATGGGCCAAGCNAGAATGTGGCTTC 817

Qy    790 tgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgttgattttgttg 849
             ||| |     | | |||  | || |||| |||  |||  | || || || | || ||
Db    818 ATAGAATCAAACTGGGATTTGGAGTTTCTCACCCAACCTTTACCAATGTGGACTATGGTG 877

Qy    850 gaggattccactggcaaacctgccaaaccccctacctaaggaaatggaggagtttgtacag 909
          ||||||    | _/ |||||||  |||||   |||||| |  ||  ||  |||| |||
Db    878 GAGGAT---GCATTGCAACCTGCTTAACCCTGGCCTAAGGGATAGGGAGA-ATTGTCCAG 933

Qy    910 agctctggagaaaatggtgttgtggtgttttctctggggtcagtgataa 958
          |  |||| | |   ||||      |  |||    |||      ||  |||
Db    934 ACTTCTGAAAACCCTGGTTGGTGTGCTTTTCTCTTGGTCAATGTTGTAA 982
```

RESULT 11
AA255225
LOCUS AA255225 616 bp mRNA EST 14-MAR-1997
DEFINITION mz85d04.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720199 5' similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR, MICROSOMAL (HUMAN); gb:X06358 Mouse mRNA for UDP-glucuronosyltransferase (MOUSE);, mRNA sequence.
ACCESSION AA255225
VERSION AA255225.1 GI:1889794
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 616)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:445695
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 443.
FEATURES Location/Qualifiers
source 1. .616
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:720199"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 179 a 124 c 140 g 173 t
ORIGIN

Query Match 23.0%; Score 394; DB 4; Length 616;

```
  Best Local Similarity   79.4%;  Pred. No. 1.4e-92;
  Matches  491;  Conservative    0;  Mismatches  125;  Indels    2;  Gaps    2;

Qy     537 tttactcctggctacacaattgaaaggcacagtggaggactgatttttccctccttcctac 596
           ||  ||||||||||||    ||||||   ||||||||||||   | |||||||| || ||
Db       1 TTCTCTCCTGGCTACTATCTTGAAAAATACAGTGGAGGACTCCCTCTCCCTCCCTCTTAT 60

Qy     597 atacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaat 656
           |||||  |   ||  |||||  |||||||| |||||||| ||||  ||||||||||| |||
Db      61 GTACCTGTAATTTTGTCAGGATTAAGTGGTCAAATGACATTCAAAGAGAGGGTAAAGAAT 120

Qy     657 atgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgggat 716
           |||||  | | ||||||||||||||||||||||||  |  | |     ||||||  ||||||
Db     121 ATGATATGTATGCTTTATTTTGACTTTTGGTTCCAGACATTTAGGGAGAAGAAATGGGAT 180

Qy     717 cagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagct 776
           ||||||||||  |||  ||| ||||| || || ||||||  ||||||||||||| |||||
Db     181 CAGTTTTACAGCGAAACTTTGGGAAGGCCAACCACCTTAATTGAGACAATGGGCAAAGCA 240

Qy     777 gacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaac 836
           || || ||||| ||  ||  ||  |||  |||  | ||||||| |||  |||||||||
Db     241 GAAATGTGGCTCATTAGATCCTATTGGGATTTGGAGTTTCCTCACCCAACCTTACCAAAT 300

Qy     837 gttgattttgttggaggattccactggcaaacctgccaaacccctacctaaggaaatgga 896
           |||  | | ||||||||||  |||| | ||||||||||  |||||  | ||||||||||||||
Db     301 GTTTACTATGTTGGAGGAGTCCA-TTGCAAACCTGCTAAACCATTGCCTAAGGAAATGGA 359

Qy     897 ggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgat 956
           || ||||| ||||||||||||||  |||||||  |||||||||||||||||| | || |
Db     360 AGACTTTGTTCAGAGCTCTGGAGAGCATGGTGTGGTGGTGTTTTCTCTGGGGTAAATGGT 419

Qy     957 aagtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccaca 1016
           |||||  |||||||  |||||  ||||||||  |||||||    || ||||||  ||||  |||||
Db     420 CAGTAATATGACAGAAGAAAAGGCCAATGCAATTGCATGGGCTCTTGCCCAGATTCCACA 479

Qy    1017 aaaggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggct 1076
           |||||||||  ||||||||||  ||  ||  |  ||||   |||||||  | ||||||  |  |
Db     480 AAAGGTTCTTTGGAGATTTGACGGCAAAACTCCAGCCACCTTAGGACCCAATACCAGAAT 539

Qy    1077 gtataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataac 1136
           ||  ||||||  |  |||||  |||||  ||||||||  |||||||  ||  ||||||||
Db     540 CTACAAGTGGCTTCCCCAAAATGATCTTCTTGGTCATCC-AAAACCAAAGGCTTTATAAC 598

Qy    1137 tcatggtggagccaatgg 1154
           |||||||||  ||  |||||
Db     599 TCATGGTGGCGCGAATGG 616


RESULT  12
AA987868/c
LOCUS       AA987868      482 bp    mRNA            EST       02-JUN-1998
DEFINITION  oq81f04.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1592767 3'
            similar to gb:J05428 UDP-GLUCURONOSYLTRANSFERASE 2B7 PRECURSOR,
            MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION   AA987868
VERSION     AA987868.1  GI:3173232
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1592767"
                     /clone_lib="NCI_CGAP_Kid6"
                     /sex="mixed"
                     /tissue_type="kidney tumor"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3' Average insert size:  1.0 kb."
BASE COUNT      136 a    104 c     82 g    160 t
ORIGIN


  Query Match             22.9%;  Score 392.4;  DB 14;  Length 482;
  Best Local Similarity   92.2%;  Pred. No. 3.4e-92;
  Matches  426;  Conservative    0;  Mismatches   31;  Indels    5;  Gaps    1;

Qy    1252 gctgttagattggacttcaacacaatgtcgagtacagacctgctgaatgcactgaagaca 1311
           |||||||||  | |||||||||||||||||||||||||||| |||||||||| |||||| |
Db     482 GCTGTTAGAGGGTACTTCAACACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGA 423

Qy    1312 gtaattaatgatcctttatataaagagaatattatgaaattatcaagaattcaacatgat 1371
           |||||||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db     422 GTAATTAATGATCCTTCATATAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGAT 363

Qy    1372 caaccagtaaagcccctggatcgagcagtcttctggattgaatttgtcatgccccacaaa 1431
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     362 CAACCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAA 303

Qy    1432 ggagccaaacaccttcgagttgcagcccatgacctcacctggttccagtaccactctttg 1491
           ||||| |||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db     302 GGAGCTAAACACCTTCGGGTTGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTG 243

Qy    1492 gatgtgattgggtttctgctggcctgtgtggcaactgtgatatttatcatcacaaagttt 1551
           |||||||||||||| ||||||| |||||||||||||||||||||||| ||||||| | |
Db     242 GATGTGATTGGGTTCCTGCTGGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGT 183

Qy    1552 tgtctgttttgtttctggaagtttgctagaaaagggaagaagggaaaaagagattagtta 1611
           |||||||||||||||||||||||||||||||||||||  ||||||||||||  |||||||||
Db     182 TGTCTGTTTTGTTTCTGGAAGTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTA 123

Qy    1612 tgtctgacatttgaagctggaaaaccagatagataggacaacttcagtttattccagcaa 1671
           | |||||  |||||||||||||||||  |||||  |  |||  ||||||||||||||||||||
Db     122 TATCTGAGATTTGAAGCTGGAAAACCTGATAGGTGAGACTACTTCAGTTTATTCCAGCAA 63

Qy    1672 gaaagaaaagattgttatgcaagatttctttcttcctgtgac 1713
           |     ||||||||| ||||||||||||||||||||||| |||
Db      62 G-----AAAGATTGTGATGCAAGATTTCTTTCTTCCTGAGAC 26


RESULT  13
AW604423/c
LOCUS       AW604423      517 bp    mRNA            EST       23-MAR-2000
DEFINITION  RC3-CT0254-200100-024-g07 CT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW604423
VERSION     AW604423.1  GI:7309177
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
```

```
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-CT0254-
            200100-024-g07&t3=2000-01-20&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 517.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0254"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      170 a    119 c     91 g    137 t
ORIGIN
```

```
  Query Match             22.6%;  Score 386.6;  DB 117;  Length 517;
  Best Local Similarity   86.2%;  Pred. No. 1.1e-90;
  Matches  439;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy     474 gtttttccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctc 533
           |||  |||||||||||||||||||||||| ||||||||||||||  ||| ||||||||||
Db     516 GTTAATCCCTGTGGTGAGCTGCTGGCTGAGCTACTTAACATACCCTTTCTGTACAGTCTC 457

Qy     534 cgctttactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcc 593
           |||||  ||  |||||||||| |||| | | |  |||||||| |  | ||||||||||||
Db     456 CGCTTCTCTGTTGGCTACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCC 397

Qy     594 tacatacctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaa 653
           ||  ||||| |||||||||| |||||||||||||||||| ||||||||||||  |||||
Db     396 TATGTACCTGTTGTTATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAA 337

Qy     654 aatatgatctatgtgctttattttgacttttggttccaaatgtctgatatgaagaagtgg 713
           ||||||||  ||| ||||||||||||||||||||||  |||    | |||| ||||||||||
Db     336 AATATGATATATATGCTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGG 277

Qy     714 gatcagttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaa 773
           || |||||||| |||||||| |||||||||||||||| ||||||||||||||||||  |||
Db     276 GACCAGTTTTATAGTGAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAA 217

Qy     774 gctgacatatggcttatgcgaaactcctggagttttcaatttcctcatccattcttacca 833
           ||||| || ||||| || |||| ||  ||   |||| |||||||||  ||||||||||
Db     216 GCTGAAATGTGGCTCATTCGAACCTATTGTGATTTTGAATTTCCTCGCCCATTCTTACCA 157

Qy     834 aacgttgattttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaat 893
           || |||||||||||||||||  |||| | ||||| |||||||||  | ||| |||||||
Db     156 AATGTTGATTTTGTTGGAGGACTTCACT-GTAAACCAGCCAAACCCTTGCCTTAGGAAAT 98

Qy     894 ggaggagtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagt 953
           ||| |||||||| |||||||||||||||||||||| ||||||||||||||||||||||  |
Db      97 GGAAGAGTTTGTGCAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGAT 38

Qy     954 gataagtaacatgacagcagaaagggcca 982
           ||| ||||||||| ||| |||||| ||||
```

```
Db      37 GATCAGTAACATGTCAGAAGAAAGTGCCA 9
```

```
RESULT  14
BF786091
LOCUS       BF786091      792 bp    mRNA            EST       12-JAN-2001
DEFINITION  602110662F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4239110
            5', mRNA sequence.
ACCESSION   BF786091
VERSION     BF786091.1  GI:12091127
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 792)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9851  row: k  column: 15
            High quality sequence start: 3
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1. .792
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4239110"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      216 a    165 c    186 g    225 t
ORIGIN
```

```
  Query Match             22.5%;  Score 385;  DB 169;  Length 792;
  Best Local Similarity   74.8%;  Pred. No. 3.4e-90;
  Matches  535;  Conservative    0;  Mismatches  175;  Indels    5;  Gaps    4;

Qy     479 tccctgtggtgagctgctggctgcgctacttaacatacggtttgtgtacagtctccgctt 538
           | ||||||| |||||  | ||||  || ||  | || |  ||| |||||||| |||||||
Db      41 TTCCTGTGGGGAGCTCATAGCTGAACTGCTCCAGATTCCTTTTCTGTACAGTATCCGCTT 100

Qy     539 tactcctggctacacaattgaaaggcacagtggaggactgattttccctccttcctacat 598
             |||| ||||||  |||||||| |  ||| ||  || |  |  ||||||| || ||  |
Db     101 CTCTCCAGGCTACCAAATTGAAAAGTCCAGCGGGAGATTTCTACTCCCTCCCTCTTATGT 160

Qy     599 acctattgttatgtcaaaattaagtgatcaaatgactttcatggagagggtaaaaaatat 658
           |||| |  || |||||  |||| |||  |||||||| ||||| ||||| ||||||||||
Db     161 ACCTGTAATTCTGTCAGGATTAGGTGGACAAATGACATTCATAGAGAGAGTAAAAAATAT 220

Qy     659 gatctatgtgctttatttttgacttttggttccaaatgtctgatatgaagaagtgggatca 718
           ||| |    |||||||||     |||||||||| |||| | ||   ||||| |||||
Db     221 GATATGCAGGCTTTATTT--GACTTTGGTTCCAGATGTTTAATGATAAGAAATGGGACTC 278

Qy     719 gttttacagtgaagttttaggaagacccactaccttatttgagacaatgggaaaagctga 778
           ||||||||||||  ||| ||||| ||||| ||||||  ||||||||||||| || || ||
Db     279 ATTTTACAGTGAATATTTGGGAAGGCCCACCACCTTAGCTGAGACAATGGGCAAGGCAGA 338

Qy     779 catatggcttatgcgaaactcctggagttttcaatttcctcatccattcttaccaaacgt 838
```

```
          || ||||| ||  ||  |  |||| ||| | |||||||| ||| |||||||||| ||
Db    339 AATGTGGCTCATTAGATCCAACTGGGATTTGGAGTTTCCTCACCCAACCTTACCAAATGT 398

Qy    839 tgattttgttggaggattccactggcaaacctgccaaaccccctacctaaggaaatggagg 898
          ||| | ||||||||||| |||||| |||||||||| |||||  | |||||||| ||||| |
Db    399 TGACTATGTTGGAGGACTCCACT-GCAAACCTGCTAAACCTTTGCCTAAGGATATGGAAG 457

Qy    899 agtttgtacagagctctggagaaaatggtgttgtggtgttttctctggggtcagtgataa 958
          | ||||| ||||||||||||||  ||||||| |||||||||| || |||||| || | |
Db    458 AATTTGTCCAGAGCTCTGGAGACCATGGTGTGGTGGTGTTTTCACTTGGGTCAATGGTTA 517

Qy    959 gtaacatgacagcagaaagggccaatgtaattgcaacagcccttgccaagatcccacaaa 1018
          |||||||||||| |||||  |||||   |||||||   |||||||| |||| |||||||
Db    518 GTAACATGACAGAAGAAAAAGCCAACACAATTGCATGGGCCCTTGCCCAGATTCCACAAA 577

Qy   1019 aggttctgtggagatttgatgggaataaaccagatgccttaggtctcaatactcggctgt 1078
          ||||||| |||| |||||||||| || |  ||||   | ||||| | ||||||  |  |||
Db    578 AGGTTCT-TGGAAATTTGATGGCAAAACCCCAGCAACTTTAGGACACAATACGAGAGTGT 636

Qy   1079 ataagtggataccccagaatgaccttctaggtcatccaaaaaccagagcttttataactc 1138
          | |||||| | |||||||||||||| || |||||||||||| || |    ||   ||||
Db    637 ACAAGTGG-TCCCCCAGAATGACCTCCTTGGTCATCCAAAAAACAAAAGCCTTTGTACTC 695

Qy   1139 atggtggagccaatggcatctatgaggcaatctaccatgggatccctatggtggg 1193
          ||||||| |||| |          ||  ||||| ||||||||||     ||||
Db    696 ATGGTGGGGCCATGGTGTCTATGAGGGTGATCTATCATGGGATCCTATGATTGGG 750


RESULT 15
BE013295
LOCUS       BE013295      554 bp    mRNA            EST       09-JUL-2000
DEFINITION  123199 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE013295
VERSION     BE013295.1  GI:8274259
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 50  row: G  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT      150 a    120 c    124 g    160 t
ORIGIN
```

```
  Query Match             22.3%;  Score 382.6;  DB 161;  Length 554;
  Best Local Similarity   81.7%;  Pred. No. 1.3e-89;
  Matches  442;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy   1039 gggaataaaccagatgccttaggtctcaatactcggctgtataagtggataccccagaat 1098
          ||||| |||||||| | |||||  ||||||||||||||| | ||||| || ||||||
Db     14 GGGAAGAAACCAGAAACATTAGGAGCCAATACTCGGCTGTATGAATGGATTCCACAGAAT 73

Qy   1099 gaccttctaggtcatccaaaaaccagagcttttataactcatggtggagccaatggcatc 1158
          || ||||| |||||||| ||||||||||||||| |||||  ||||| |||||||| |||
Db     74 GATCTTCTTGGTCATCCCCAAACCAGAGCTTTTATTACTCACTGTGGAACCAATGGGATC 133

Qy   1159 tatgaggcaatctaccatgggatccctatggtgggcattccattgttttgggatcaacct 1218
          ||||| || || |||||||||||||||||||||| |||||  ||||| | ||||| | |
Db    134 TATGAAGCTATTTACCATGGGATCCCTATGGTGGGAATTCCCATGTTTGGTGATCAGCAT 193

Qy   1219 gataacattgctcacatgaaggccaagggagcagctgttagattggacttcaacacaatg 1278
          ||||| |||||||   | || || || || |||||||||  |||  ||||  ||||||||
Db    194 GATAATATTGCTCGTCTAAAAGCTAAAGGGGCAGCTGTTGAATTAAACTTGCACACAATG 253

Qy   1279 tcgagtacagacctgctgaatgcactgaagacagtaattaatgatcctttatataaagag 1338
          ||||| | || ||||| |||||  || || |||| |||||  | ||||  ||||||||
Db    254 ACGAGTTCTGATCTGCTTAATGCTTTGGAGGCAGTTATTAACAACCCTTCCTATAAAGAA 313

Qy   1339 aatattatgaaattatcaagaattcaacatgatcaaccagtaaagcccctggatcgagca 1398
          |||  |||||| |||||||||||||| ||||||||||  |||||||||||| ||||||
Db    314 AATGCTATGAAGTTATCAAGAATTCACCATGATCAACCTATAAAGCCCCTGGACCGAGCA 373

Qy   1399 gtcttctggattgaatttgtcatgccccacaaaggagccaaacaccttcgagttgcagcc 1458
          |||||||||||||| |||||||||| |||||||||||||| ||||| ||    ||  ||
Db    374 GTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGGCCAGCTTCC 433

Qy   1459 catgacctcacctggttccagtaccactctttggatgtgattgggtttctgctggcctgt 1518
          || |||||||||||| |||||||||||| |||||||||||||||| |||||||| |||
Db    434 CACGACCTCACCTGGTACCAGTACCACTCTCTGGATGTGATTGGGTTCCTGCTGGCTTGT 493

Qy   1519 gtggcaactgtgatatttatcatcacaaagttttgtctgttttgtttctggaagtttgct 1578
          ||||||||| | |||||| |  |||  || | ||  ||||||||||  |  ||||||| |
Db    494 GTGGCAACTATTATATTTCTGGTCATTAAATGTTTCCTGTTTTGTTGTTCAAAGTTTGGT 553

Qy   1579 a 1579
          |
Db    554 A 554


Search completed: April 30, 2001, 12:17:28
Job time: 12517 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:15:44 ; Search time 22667.1 Seconds
(without alignments)
2718.307 Million cell updates/sec

Title: US-09-721-183-20
Perfect score: 10006
Sequence: 1 ttcctccgcgaaggctcctt.....agttttccaccagttacaca 10006

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		9951	99.5	10011	88	AF183810	AF183810 Homo sapi
2	c	6476.6	64.7	159505	68	AC023142	AC023142 Homo sapi
3		6099.8	61.0	167123	63	AC013815	AC013815 Homo sapi
4		5271	52.7	5507	88	AF264784	AF264784 Homo sapi
5		2474.4	24.7	2596	88	AK000948	AK000948 Homo sapi
6		1568.6	15.7	1587	89	AK021511	AK021511 Homo sapi
7		1525	15.2	1551	89	AK021470	AK021470 Homo sapi
8	c	1127.4	11.3	310002	88	AF178030	AF178030 Homo sapi
9		931.4	9.3	996	10	AX067304	AX067304 Sequence
10	c	827.2	8.3	167123	63	AC013815	AC013815 Homo sapi
11		668.8	6.7	674	10	AX067327	AX067327 Sequence

c	12	460.8	4.6	350000	88	AF130342	AF130342 Homo sapi
c	13	310	3.1	50698	77	AC084171	AC084171 Homo sapi
c	14	258.4	2.6	67407	63	AC013789	AC013789 Homo sapi
	15	93.4	0.9	67407	63	AC013789	AC013789 Homo sapi
c	16	89.4	0.9	224448	84	PFMAL4P4	AL035477 Plasmodiu
c	17	87.2	0.9	195969	76	AC074354	AC074354 Oryza sat
	18	82.6	0.8	879	53	CNS01JRG	AL147405 Anopheles
	19	81.8	0.8	53932	68	AC023371	AC023371 Homo sapi
c	20	80.4	0.8	936	54	CNS07ANM	AL436808 T7 end of
	21	79.8	0.8	1084	53	CNS06PLB	AL409509 T7 end of
c	22	79.4	0.8	130540	76	AC079417	AC079417 Mus muscu
	23	78.4	0.8	1169	53	CNS06KHQ	AL402900 T3 end of
	24	78	0.8	897	54	CNS07ABZ	AL436389 T7 end of
c	25	77.8	0.8	22243	6	PFAVAR23A	L40609 Plasmodium
	26	77.4	0.8	83440	69	AC024285	AC024285 Homo sapi
c	27	77	0.8	192929	60	AC005505	AC005505 Plasmodiu
c	28	77	0.8	256172	60	AC005139	AC005139 Plasmodiu
c	29	77	0.8	310779	60	AC005140	AC005140 Plasmodiu
	30	76.6	0.8	187279	70	AC026698	AC026698 Homo sapi
	31	76.4	0.8	184535	83	CNS05TCJ	AL355100 Homo sapi
	32	76.2	0.8	99003	81	AL390756	AL390756 Homo sapi
c	33	76.2	0.8	127709	83	CNS05TC1	AL352983 Homo sapi
	34	76	0.8	115218	91	HS159A1	AL034397 Human DNA
	35	75.6	0.8	90011	90	AL360007	AL360007 Human DNA
	36	75.2	0.8	321003	84	PFMAL4P3	AL035476 Plasmodiu
c	37	75	0.7	161891	60	AC008206	AC008206 Drosophil
	38	74.8	0.7	12600	63	AC013952	AC013952 Drosophil
	39	74.8	0.7	156550	64	AC015830	AC015830 Homo sapi
	40	74.8	0.7	256172	60	AC005139	AC005139 Plasmodiu
c	41	74.6	0.7	759	53	CNS06QXV	AL411257 T7 end of
	42	74.6	0.7	840	14	CNS018QK	AL110675 Botrytis
	43	74.6	0.7	60604	68	AC023466	AC023466 Homo sapi
	44	74.6	0.7	143585	63	AC013349	AC013349 Homo sapi
	45	74	0.7	1167	53	CNS07360	AL427102 clone BA0

ALIGNMENTS

```
RESULT   1
AF183810
LOCUS       AF183810    10011 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens zinc finger transcription factor TRPS1 mRNA, complete
            cds.
ACCESSION   AF183810
VERSION     AF183810.1  GI:6684533
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10011)
  AUTHORS   Momeni,P., Glockner,G., Schmidt,O., von Holtum,D., Albrecht,B.,
            Gillessen-Kaesbach,G., Hennekam,R., Meinecke,P., Zabel,B.,
            Rosenthal,A., Horsthemke,B. and Ludecke,H.J.
  TITLE     Mutations in a new gene, encoding a zinc-finger protein, cause
            tricho-rhino-phalangeal syndrome type I
  JOURNAL   Nat. Genet. 24 (1), 71-74 (2000)
  MEDLINE   20082812
REFERENCE   2  (bases 1 to 10011)
  AUTHORS   Momeni,P., Gloeckner,G., Schmidt,O., von Holtum,D., Albrecht,B.,
            Gillessen-Kaesbach,G., Hennekam,R., Meinecke,P., Zabel,B.,
            Rosenthal,A., Horsthemke,B. and Luedecke,H.-J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1999) Institut fuer Humangenetik,
            Universitaetsklinikum, Hufelandstr. 55, D-45122 Essen, Germany
FEATURES             Location/Qualifiers
     source          1. .10011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24.1"
     CDS             639. .4484
                     /note="tricho-rhino-phalangeal syndrome type I protein"
                     /codon_start=1
                     /product="zinc finger transcription factor TRPS1"
                     /protein_id="AAF23614.1"
                     /db_xref="GI:6684534"
                     /translation="MVRKKNPPLRNVASEGEGQILEPIGTESKVSGKNKEFSADQMSE
                     NTDQSDAAELNHKEEHSLHVQDPSSSSKKDLKSAVLSEKAGFNYESPSKGGNFPSFPH
                     DEVTDRNMLAFSFPAAGGVCEPLKSPQRAEADDPQDMACTPSGDSLETKEDQKMSPKA
                     TEETGQAQSGQANCQGLSPVSVASKNPQVPSDGGVRLNKSKTDLLVNDNPDPAPLSPE
                     LQDFKCNICGYGYYGNDPTDLIKHFRKYHLGLHNRTRQDAELDSKILALHNMVQFSHS
                     KDFQKVNRSVFSGVLQDINSSRPVLLNGTYDVQVTSGGTFIGIGRKTPDCQGNTKYFR
                     CKFCNFTYMGNSSTELEQHFLQTHPNKIKASLPSSEVAKPSEKNSNKSIPALQSSDSG
                     DLGKWQDKITVKAGDDTPVGYSVPIKPLDSSRQNGTEATSYYWCKFCSFSCESSSSLK
                     LLEHYGKQHGAVQSGGLNPELNDKLSRGSVINQNDLAKSSEGETMTKTDKSSSGAKKK
                     DFSSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLHNIHKCTIKHCPF
                     CPRGLCSPEKHLGEITYPFACRKSNCSHCALLLLHLSPGAAGSSRVKHQCHQCSFTTP
                     DVDVLLFHYESVHESQASDVKQEANHLQGSDGQQSVKESKEHSCTKCDFITQVEEEIS
                     RHYRRAHSCYKCRQCSFTAADTQSLLEHFNTVHCQEQDITTANGEEDGHAISTIKEEP
                     KIDFRVYNLLTPDSKMGEPVSESVVKREKLEEKDGLKEKVWTESSSDDLRNVTWRGAD
                     ILRGSPSYTQASLGLLTPVSGTQEQTKTLRDSPNVEAAHLARPIYGLAVETKGFLQGA
                     PAGGEKSGALPQQYPASGENKSKDESQSLLRRRRGSGVFCANCLTTKTSLWRKNANGG
                     YVCNACGLYQKLHSTPRPLNIIKQNNGEQIIRRRTRKRLNPEALQAEQLNKQQRGSNE
                     EQVNGSPLERRSEDHLTESHQREIPLPSLSKYEAQGSLTKSHSAQQPVLVSQTLDIHK
                     RMQPLHIQIKSPQESTGDPGNSSSVSEGKGSSERGSPIEKYMRPAKHPNYSPPGSPIE
                     KYQYPLFGLPFVHNDFQSEADWLRFWSKYKLSVPGNPHYLSHVPGLPNPCQNYVPYPT
                     FNLPPHFSAVGSDNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVKTEK
                     VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHGDSGPFQCSICQHLCTDKYDFTTHI
                     QRGLHRNNAQVEKNGKPKE"
BASE COUNT     3067 a    1963 c    2056 g    2925 t
ORIGIN
```

```
  Query Match             99.5%;  Score 9951;  DB 88;  Length 10011;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 10006;  Conservative    0;  Mismatches    0;  Indels    5;  Gaps    5;

Qy       1 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTCCTCCGCGAAGGCTCCTTTGATATTAATAGTGTTGGTGTCTTGAAACTGACGTAATGC 60

Qy      61 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCGGAGACTGAGGTCCTGACAAGCGATAACATTTCTGATAAAGACCCGATCTTACTGCAA 120

Qy     121 tctctagcgtcctcttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCTCTAGCGTCCTCTTTTTTGGTGCTGCTGGTTTCTCCAGACCTCGCGTCCTCTCGATTG 180

Qy     181 ctctctcgccttcctatttctttttttttttttaaacaaaaaacaacacccccctcccct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTCTCTCGCCTTCCTATTTCTTTTTTTTTTTTTAAACAAAAAACAACACCCCCTCCCCT 240

Qy     241 ctcccacccggcaccgggcacatccttgctctatttccttttctctttctctctctctc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCCCACCCGGCACCGGGCACATCCTTGCTCTATTTCCTTTCTCTTTCTCTCTCTCTC 300

Qy     301 tctctttttaataagggtgggggagggaaaggggggggaggcaggaaagacctttttct 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCTCTTTTTTAATAAGGGTGGGGGAGGGAAAGGGGGGGGAGGCAGGAAAGACCTTTTTCT 360

Qy     361 ctccccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcatggcttt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTCCCCCCCGCAATAATCCAAGATCAACTCTGCAAACAACAGAAGACGGTTCATGGCTTT 420

Qy     421 ggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatcaacagat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGCCGCCGCGCCACCATCTTTCGGGCTGCCGAGGGTGTTCTTGACGATTAATCAACAGAT 480

Qy     481 gtacagatcagctctcaaaatgtcttctgtgtcttctgagcgtcttctaagacaattgca 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTACAGATCAGCTCTCAAAATGTCTTCTGTGTCTTCTGAGCGTCTTCTAAGACAATTGCA 540

Qy     541 ttagcctcctgctagttgactaatagaattaataattgtaaaaagcactctaaagccaca 600
```

```
Db    541 TTAGCCTCCTGCTAGTTGACTAATAGAATTAATAATTGTAAAAAGCACTCTAAAGCCACA 600
Qy    601 tgccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaagaaccccc 660
Db    601 TGCCTTATGAAGTCAATGCTGGGTATGATTTTACAAATATGGTCCGGAAAAAGAACCCCC 660
Qy    661 ctctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggtacagaaa 720
Db    661 CTCTGAGAAACGTTGCAAGTGAAGGCGAGGGCCAGATCCTGGAGCCTATAGGTACAGAAA 720
Qy    721 gcaaggtatctggaaagaacaaagaattctctgcagatcagatgtcagaaaatacggatc 780
Db    721 GCAAGGTATCTGGAAAGAACAAAGAATTCTCTGCAGATCAGATGTCAGAAAATACGGATC 780
Qy    781 agagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaagatccat 840
Db    781 AGAGTGATGCTGCAGAACTAAATCATAAGGAGGAACATAGCTTGCATGTTCAAGATCCAT 840
Qy    841 cttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggcttcaatt 900
Db    841 CTTCTAGCAGTAAGAAGGACTTGAAAAGCGCAGTTCTGAGTGAGAAGGCTGGCTTCAATT 900
Qy    901 atgaaagccccagtaagggaggaaactttccctcctttccgcatgatgaggtgacagaca 960
Db    901 ATGAAAGCCCCAGTAAGGGAGGAAACTTTCCCTCCTTTCCGCATGATGAGGTGACAGACA 960
Qy    961 gaaatatgttggctttctcatttccagctgctgggggagtctgtgagcccttgaagtctc 1020
Db    961 GAAATATGTTGGCTTTCTCATTTCCAGCTGCTGGGGGAGTCTGTGAGCCCTTGAAGTCTC 1020
Qy   1021 cgcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcaggggactcac 1080
Db   1021 CGCAAAGAGCAGAGGCAGATGACCCTCAAGATATGGCCTGCACCCCCTCAGGGGACTCAC 1080
Qy   1081 tggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggcaagcac 1140
Db   1081 TGGAGACAAAGGAAGATCAGAAGATGTCACCAAAGGCTACAGAGGAAACAGGGCAAGCAC 1140
Qy   1141 agagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaaacccac 1200
Db   1141 AGAGTGGTCAAGCCAATTGTCAAGGTTTGAGCCCAGTTTCAGTGGCCTCAAAAAACCCAC 1200
Qy   1201 aagtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactggtgaatg 1260
Db   1201 AAGTGCCTTCAGATGGGGGTGTAAGACTGAATAAATCCAAAACTGACTTACTGGTGAATG 1260
Qy   1261 acaacccagacccggcacctctgtctccagagcttcaggactttaaatgcaatatctgtg 1320
Db   1261 ACAACCCAGACCCGGCACCTCTGTCTCCAGAGCTTCAGGACTTTAAATGCAATATCTGTG 1320
Qy   1321 gatatggttactacggcaacgaccccacagatctgattaagcacttccgaaagtatcact 1380
Db   1321 GATATGGTTACTACGGCAACGACCCCACAGATCTGATTAAGCACTTCCGAAAGTATCACT 1380
Qy   1381 taggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttggcccttc 1440
Db   1381 TAGGACTGCATAACCGCACCAGGCAAGATGCTGAGCTGGACAGCAAAATCTTGGCCCTTC 1440
Qy   1441 ataacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctgtgtttt 1500
Db   1441 ATAACATGGTGCAGTTCAGCCATTCCAAAGACTTCCAGAAGGTCAACCGTTCTGTGTTTT 1500
Qy   1501 ctggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacctatgatg 1560
Db   1501 CTGGTGTGCTGCAGGACATCAATTCTTCAAGGCCTGTTTTACTAAATGGGACCTATGATG 1560
Qy   1561 tgcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagattgccaag 1620
Db   1561 TGCAGGTGACTTCAGGTGGAACATTCATTGGCATTGGACGGAAAACACCAGATTGCCAAG 1620
Qy   1621 ggaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaactcatcca 1680
Db   1621 GGAACACCAAGTATTTCCGCTGTAAATTCTGCAATTTCACTTATATGGGCAACTCATCCA 1680
Qy   1681 ccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctctcccct 1740
Db   1681 CCGAATTAGAACAACATTTTCTTCAGACTCACCCAAACAAAATAAAAGCTTCTCTCCCCT 1740
Qy   1741 cctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcacttcaat 1800
Db   1741 CCTCTGAGGTTGCAAAACCTTCAGAGAAAAACTCTAACAAGTCCATCCCTGCACTTCAAT 1800
Qy   1801 ccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcaggagatg 1860
Db   1801 CCAGTGATTCTGGAGACTTGGGAAAATGGCAGGACAAGATAACAGTCAAAGCAGGAGATG 1860
Qy   1861 acactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaaa-ggta 1919
Db   1861 ACACTCCTGTTGGGTACTCAGTGCCCATAAAGCCCCTCGATTCCTCTAGACAAAATGGTA 1920
Qy   1920 cagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcatctagct 1979
Db   1921 CAGAGGCCACCAGTTACTACTGGTGTAAATTTTGTAGTTTCAGCTGTGAGTCATCTAGCT 1980
Qy   1980 cacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcggcctta 2039
Db   1981 CACTTAAACTGCTAGAACATTATGGCAAGCAGCACGGAGCAGTGCAGTCAGGCGGCCTTA 2040
Qy   2040 atccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatctagcca 2099
Db   2041 ATCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCATTAATCAGAATGATCTAGCCA 2100
Qy   2100 aaagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggctaaaaaga 2159
Db   2101 AAAGTTCAGAAGGAGAGACAATGACCAAGACAGACAAGAGCTCGAGTGGGGCTAAAAAGA 2160
Qy   2160 aggacttctccagcaagggagccgaggataatatggtaacgagctataattgtcagttct 2219
Db   2161 AGGACTTCTCCAGCAAGGGAGCCGAGGATAATATGGTAACGAGCTATAATTGTCAGTTCT 2220
Qy   2220 gtgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccacttctcc 2279
Db   2221 GTGACTTCCGATATTCCAAAAGCCATGGCCCTGATGTAATTGTAGTGGGGCCACTTCTCC 2280
Qy   2280 gtcattatcaacagctccataacattcacaagtgtaccattaaacactgtccattctgtc 2339
Db   2281 GTCATTATCAACAGCTCCATAACATTCACAAGTGTACCATTAAACACTGTCCATTCTGTC 2340
Qy   2340 ccagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttgcttgta 2399
Db   2341 CCAGAGGACTTTGCAGCCCAGAAAAGCACCTTGGAGAAATTACTTATCCGTTTGCTTGTA 2400
Qy   2400 gaaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctggggcggctg 2459
Db   2401 GAAAAAGTAATTGTTCCCACTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGCGGCTG 2460
Qy   2460 gaagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacgtagatg 2519
Db   2461 GAAGCTCGCGAGTCAAACATCAGTGCCATCAGTGTTCATTCACCACCCCTGACGTAGATG 2520
Qy   2520 tactcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaacaagaag 2579
Db   2521 TACTCCTCTTTCACTATGAAAGTGTGCATGAGTCCCAAGCATCGGATGTCAAACAAGAAG 2580
Qy   2580 caaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaacactcat 2639
Db   2581 CAAATCACCTGCAAGGATCGGATGGGCAGCAGTCTGTCAAGGAAAGCAAAGAACACTCAT 2640
Qy   2640 gtaccaaatgtgattttattacccaagtggaagaagagatttcccgacactacaggagag 2699
Db   2641 GTACCAAATGTGATTTTATTACCCAAGTGGAAGAAGAGATTTCCCGACACTACAGGAGAG 2700
Qy   2700 cacacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagtcactac 2759
Db   2701 CACACAGCTGCTACAAATGCCGTCAGTGCAGTTTTACAGCTGCCGATACTCAGTCACTAC 2760
```

```
Qy   2760 tggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacggcgaag 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGGAGCACTTCAACACTGTTCACTGCCAGGAACAGGACATCACTACAGCCAACGGCGAAG 2820

Qy   2820 aggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcagggtctaca 2879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 AGGACGGTCATGCCATATCCACCATCAAAGAGGAGCCCAAAATTGACTTCAGGGTCTACA 2880

Qy   2880 atctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtgaagagag 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 ATCTGCTAACTCCAGACTCTAAAATGGGAGAGCCAGTTTCTGAGAGTGTGGTGAAGAGAG 2940

Qy   2940 agaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttccagtgatg 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 AGAAGCTGGAAGAGAAGGACGGGCTCAAAGAGAAAGTTTGGACCGAGAGTTCCAGTGATG 3000

Qy   3000 accttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcatacaccc 3059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACCTTCGCAATGTGACTTGGAGAGGGGCAGACATCCTGCGGGGGAGTCCGTCATACACCC 3060

Qy   3060 aagcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaagactctaa 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AAGCAAGCCTGGGGCTGCTGACGCCTGTGTCTGGCACCCAAGAGCAGACAAAGACTCTAA 3120

Qy   3120 gggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttggctgtgg 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GGGATAGTCCCAATGTGGAGGCCGCCCATCTGGCGCGACCTATTTATGGCTTGGCTGTGG 3180

Qy   3180 aaaccaagggattcctgcaggggggcgccagctggcggagagaagtctggggccctccccc 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AAACCAAGGGATTCCTGCAGGGGGCGCCAGCTGGCGGAGAGAAGTCTGGGGCCCTCCCCC 3240

Qy   3240 agcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgttacgga 3299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 AGCAGTATCCTGCATCGGGAGAAAACAAGTCCAAGGATGAATCCCAGTCCCTGTTACGGA 3300

Qy   3300 ggcgtagaggctccggtgtttttttgtgccaattgcctgaccacaaagacctctctctggc 3359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 GGCGTAGAGGCTCCGGTGTTTTTTGTGCCAATTGCCTGACCACAAAGACCTCTCTCTGGC 3360

Qy   3360 gaaagaatgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagcttcact 3419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 GAAAGAATGCAAATGGCGGATATGTATGCAACGCGTGTGGCCTCTACCAGAAGCTTCACT 3420

Qy   3420 cgactcccaggccttaaacatcattaaacaaaacaacggtgagcagattattaggagga 3479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CGACTCCCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGA 3480

Qy   3480 gaacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcaga 3539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GAACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGA 3540

Qy   3540 ggggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatc 3599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GGGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATC 3600

Qy   3600 taactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccagg 3659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 TAACTGAAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGG 3660

Qy   3660 gttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggata 3719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GTTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATA 3720

Qy   3720 ttcacaaaaggatgcaaccttttgcacattcagataaaaagtcctcaggaaagtactggag 3779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 TTCACAAAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAG 3780

Qy   3780 atccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtccta 3839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 ATCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTA 3840

Qy   3840 tagaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattg 3899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 TAGAAAAGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTG 3900

Qy   3900 aaaagtaccagtacccacttttttggacttcccttt-tacataatgacttccagagtgaag 3958
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   3901 AAAAGTACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAG 3960

Qy   3959 ctgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactact 4018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 CTGATTGGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACT 4020

Qy   4019 tgagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttca 4078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 TGAGTCACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCA 4080

Qy   4079 atctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcga 4138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 ATCTGCCTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGA 4140

Qy   4139 tcaagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcac 4198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 TCAAGCATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCAC 4200

Qy   4199 caccaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaa 4258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 CACCAAATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAA 4260

Qy   4259 gtactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaag 4318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 GTACTCAAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAG 4320

Qy   4319 tgatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatat 4378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 TGATGTATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATAT 4380

Qy   4379 gccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcata 4438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 GCCAGCATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATA 4440

Qy   4439 ggaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagcacttagc 4498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 GGAACAATGCACAAGTGGAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGC 4500

Qy   4499 acaattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaa 4558
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 ACAATTAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAA 4560

Qy   4559 gacctattaaaaaaatacttcatagagcctgccttatccaacatgaaattcccttctttt 4618
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 GACCTATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTT 4620

Qy   4619 gttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatg 4678
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 GTTATTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATG 4680

Qy   4679 agaaagaatggaagatggtaaacaatcacttttttaaaacctgttaagtcaaaaccatctt 4738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 AGAAAGAATGGAAGATGGTAAACAATCACTTTTTAAAACCTGTTAAGTCAAAACCATCTT 4740

Qy   4739 ggctaatatgtactggggaaataatccataagagatatcaccagactagaattaatatat 4798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 GGCTAATATGTACTGGGGAAATAATCCATAAGAGATATCACCAGACTAGAATTAATATAT 4800

Qy   4799 ttataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatacta 4858
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 TTATAAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGTTTACATATTATTATACTA 4860

Qy   4859 aagcagtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaa 4918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 AAGCAGTACTGGACTGGCCATTGGACCATTTGTTCCAAAACCCATAAATTGTTGCCTAAA 4920

Qy   4919 tttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaa 4978
```

```
Db  4921 TTTATAATGATCATGAAACCCTAGGCAGAGGAGGAGAAATTGAAGGTCCAGGGCAATGAA 4980
Qy  4979 agaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagattttgacc 5038
Db  4981 AGAAAAATGGCGCCCTCTCAATTTAGTCTTCTCTCATTGGCCATGTTTCAGATTTTGACC 5040
Qy  5039 tagaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtg 5098
Db  5041 TAGAAATGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTG 5100
Qy  5099 gcacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctgatatttt 5158
Db  5101 GCACGCTGTTTTTACCCAGCCCTGCCTGTACATACACATGCACACCCTCTCTGATATTTT 5160
Qy  5159 tgtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattcattttgt 5218
Db  5161 TGTCCTTTAGATGTTCAAATACTCAGTAGTCCTTTTGTTTGCGGTTTAGATTCATTTTGT 5220
Qy  5219 ccacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgatttcatttt 5278
Db  5221 CCACACATGTACCCATTTTAAAAAACAATGTCCTCGATGCTTCTGTAGTGATTTCATTTT 5280
Qy  5279 agccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaagcctcc 5338
Db  5281 AGCCAGGTATTTCTTTCTTGTGTGTGATGAACCAGTATGGATTTGCTTTTCTAAGCCTCC 5340
Qy  5339 tgttggttactaatctcacttggcacattataactaaaggaatcccctcaattcaaaagc 5398
Db  5341 TGTTGGTTACTAATCTCACTTGGCACATTATAACTAAAGGAATCCCCTCAATTCAAAAGC 5400
Qy  5399 atagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttctt 5458
Db  5401 ATAGATGGATACAAATGTCAGACCGTGGGTTTAATTTGTTTAGAACACATGGCATTTCTT 5460
Qy  5459 cacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttccaaacttt 5518
Db  5461 CACAAGGTAACCTGCTGTATTTATTTATTTTCTTTTGGTTAAATATAATTTCCAAACTTT 5520
Qy  5519 gtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgtaccagc 5578
Db  5521 GTGGTCAGGCAGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAGC 5580
Qy  5579 caatcccttcattaaatgtatacagatttagttaagtagcattaaataggattcttagaa 5638
Db  5581 CAATCCCTTCATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAA 5640
Qy  5639 gtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaagggaaa 5698
Db  5641 GTATGTCCTCATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAA 5700
Qy  5699 gctcctcagcataactgctcagggaaatagggctaaataactgaacattaaataattggt 5758
Db  5701 GCTCCTCAGCATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGGT 5760
Qy  5759 taaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactgac 5818
Db  5761 TAAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGAC 5820
Qy  5819 tttaataatttgcattatattgtcccaaccagtagtttatttttgccacggagatgtag 5878
Db  5821 TTTAATAATTTGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTAG 5880
Qy  5879 aagatattacaagctactggatgcactgtcagattaacttatttcattaaagaagttggg 5938
Db  5881 AAGATATTACAAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGG 5940
Qy  5939 agaacaaatagg-aaaaaaaaaacttatttttctagtaaatattaatgtattacatttcaa 5997
Db  5941 AGAACAAATAGGAAAAAAAAAACTTATTTTTCTAGTAAATATTAATGTATTACATTTCAA 6000
Qy  5998 ataatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaacaaagata 6057
Db  6001 ATAATGGTGCCTGACATATTGAATAATTATTTTCTACAGTGTACGTATGCAACAAAGATA 6060
Qy  6058 ttccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaaatgtag 6117
Db  6061 TTCCATCATGCATTAGAGTCAGTTCTGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAG 6120
Qy  6118 caaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgt 6177
Db  6121 CAAACAAGGTAATGAAGCAACTATTTCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGT 6180
Qy  6178 gtgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatg 6237
Db  6181 GTGCATTAAAGTTGTAAACGGTAACATGAAACAAATGAAAGTTCTTGCTATAATGGTATG 6240
Qy  6238 gaaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcactt 6297
Db  6241 GAAAACAAGAAGGAAATGAAAATATTTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTT 6300
Qy  6298 attcattccaagtacttttttttttttaatttttaagctcttaactcacattgttatgct 6357
Db  6301 ATTCATTCCAAGTACTTTTTTTTTTTTAATTTTTAAGCTCTTAACTCACATTGTTATGCT 6360
Qy  6358 taagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaaacattt 6417
Db  6361 TAAGATGATAAACATATATCCTCTTTTTATTGCTTTGTCTATGTTTCATATGAAACATTT 6420
Qy  6418 cagaaattattttgataagtgttgctggaatctgcaacgctgatttttttttgcattctg 6477
Db  6421 CAGAAATTATTTTGATAAGTGTTGCTGGAATCTGCAACGCTGATTTTTTTTTGCATTCTG 6480
Qy  6478 tagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattgttttgt 6537
Db  6481 TAGTCGCATTTGCACTCCATTTTTACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGT 6540
Qy  6538 ttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggt 6597
Db  6541 TTGGGTTTTGTTTCTTTTTCACAGTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGT 6600
Qy  6598 agagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatg 6657
Db  6601 AGAGTTCAACCAGTTCGTGACTGTTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATG 6660
Qy  6658 ttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagag 6717
Db  6661 TTGTGTTGTCATTTTCATTTTTGCATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAG 6720
Qy  6718 aaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaacgggcct 6777
Db  6721 AAAGCAAGAGACAGAAATCAGGACTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCT 6780
Qy  6778 tattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaagaagtgg 6837
Db  6781 TATTCTGATCTCACCTGTCGCGTAGCTCTAATATTCACATAAACTGAAATAAAGAAGTGG 6840
Qy  6838 aatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaattttgat 6897
Db  6841 AATGAGGAGCTTTGACATTCAAATTATGTGATGTAATTTATCTTCCTTAGGAATTTTGAT 6900
Qy  6898 ggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaaaga 6957
Db  6901 GGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAAAGA 6960
Qy  6958 gaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaatgca 7017
Db  6961 GAGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAATGCA 7020
Qy  7018 tcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacaccaga 7077
Db  7021 TCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACCAGA 7080
Qy  7078 agtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactcattaag 7137
Db  7081 AGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATTAAG 7140
```

```
Qy   7138 caatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttt 7197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 CAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTTTTT 7200

Qy   7198 attttcattcaatttttttttttctcagacagacatagtagtatcaactagcattggaa 7257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 ATTTTCATTCAATTTTTTTTTTTTCTCAGACAGACATAGTAGTATCAACTAGCATTGGAA 7260

Qy   7258 aatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaatattttt 7317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 AATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAAAATATTTTT 7320

Qy   7318 ggatagcgttgacacgatagatcttattccatacttctttattattgataatttttatttt 7377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 GGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTT 7380

Qy   7378 catttttgctttcattattatacatattttggtggagaagaggttgggctttttgaaa 7437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 CATTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTGAAA 7440

Qy   7438 gagacaaaaatttattataacactaaacactcctttttgacatattaaagcctttattc 7497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 GAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTC 7500

Qy   7498 catctctcaagatatattataaaatttatttttttaatttaagatttctgaattatttta 7557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 CATCTCTCAAGATATATTATAAAATTTATTTTTTAATTTAAGATTTCTGAATTATTTTA 7560

Qy   7558 tcttaaattgtgattttaaacgagctattatggtacggaacttttttaatgaggaattt 7617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 TCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTAATGAGGAATTT 7620

Qy   7618 catgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtg 7677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 CATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTG 7680

Qy   7678 ccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatatt 7737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 CCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATT 7740

Qy   7738 ctaaattgagctttggatcaaacttttaggccaggaccagctcatgcgttctcattcttcc 7797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7741 CTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCC 7800

Qy   7798 ttttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatagaaa 7857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7801 TTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAA 7860

Qy   7858 aatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatgactc 7917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7861 AATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTC 7920

Qy   7918 caagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg-aaaga 7976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   7921 CAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAAAGA 7980

Qy   7977 gtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaagatag 8036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7981 GTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGATAG 8040

Qy   8037 tgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagcatg 8096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8041 TGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATG 8100

Qy   8097 gggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctcttt 8156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8101 GGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTT 8160

Qy   8157 gttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctattatg 8216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8161 GTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATG 8220

Qy   8217 atccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctgagg 8276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8221 ATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGG 8280

Qy   8277 aaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattattagc 8336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8281 AAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGC 8340

Qy   8337 ctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgct 8396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8341 CTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCT 8400

Qy   8397 gccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccagtgg 8456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8401 GCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGG 8460

Qy   8457 tggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgcc 8516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8461 TGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCC 8520

Qy   8517 ttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcaaaaaa 8576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8521 TTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAAAAA 8580

Qy   8577 aaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgctgac 8636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8581 AAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGCTGAC 8640

Qy   8637 tgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaaata 8696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8641 TGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTAAATA 8700

Qy   8697 ttctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatgaga 8756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8701 TTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAATGAGA 8760

Qy   8757 agaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaagaaaga 8816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8761 AGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCATTGTTATTTTTCCCAAAGAAAGA 8820

Qy   8817 cggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagtactg 8876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8821 CGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACTG 8880

Qy   8877 tttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcatccc 8936
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8881 TTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCCC 8940

Qy   8937 agttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagtatgt 8996
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8941 AGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGT 9000

Qy   8997 cacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgtaca 9056
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9001 CACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATGTACA 9060

Qy   9057 ttcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacaccgga 9116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9061 TTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGGA 9120

Qy   9117 gaaaagcaggaataactttaccgtggaaataatgtttagctttttatcagagaaaattgtc 9176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9121 GAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTATCAGAGAAAATTGTC 9180

Qy   9177 cttctagagcatagagtcccaaaactcaattctggttttcccctgttttttttttttttt 9236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9181 CTTCTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCCTGTTTTTTTTTTTTTT 9240

Qy   9237 tttttcccaacatatgaactgcagcatatcacttttttcttttttgtgcctcaggttcctca 9296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9241 TTTTTCCCAACATATGAACTGCAGCATATCACTTTTTCTTTTTTGTGCCTCAGGTTCCTCA 9300

Qy   9297 cctgtaaaattgaaaaatatatgtattaataatattattaataataataatggtaatgta 9356
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9301 CCTGTAAAATTGAAAAATATATGTATTAATAATATTATTAATAATAATAATGGTAATGTA 9360

Qy    9357 gtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtgccaagt 9416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9361 GTACTTGTTTGTAAAGCACTTTGAGATCCTTGGTTGAAAGGCACCATAGGAGTGCCAAGT 9420

Qy    9417 attattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaaaggttg 9476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9421 ATTATTATGTGGCCAAGGGGTTATTTAAACTGTCAGTTCCCAAAGGCCAGGAAAGGTTG 9480

Qy    9477 gggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgttgacta 9536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9481 GGGTCATTTTTCTTAAAGACGAGCTGTAAATATCAACTAGGCAGCCAATAGTGTTGACTA 9540

Qy    9537 tgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctaccaataa 9596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9541 TGAAGATGCAAAACTATTACTAGGCTGATAAAATCATAGTTTCTTAATGGCTACCAATAA 9600

Qy    9597 ggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaaccacatg 9656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9601 GGCAAATATCACAATAATAAACGCCAAATTCCTTAGGGCGGACTATTTGACAACCACATG 9660

Qy    9657 gaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaatttaact 9716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9661 GAAAACTTTGGGGGAGGCATGAGGGGGGAACATCTCAAAATGCCAATGTAAAATTTAACT 9720

Qy    9717 tacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttgctaaga 9776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9721 TACAGCAATATTCACCAGCAGAAAATGTCTTTCATATGGAATGATTTCATGTTGCTAAGA 9780

Qy    9777 aaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatagttctgt 9836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9781 AAAAGAATTCAATTTGTAGTCCTGATTTGAATACTAGAATGTTGGCTATAATAGTTCTGT 9840

Qy    9837 tcttacaacacatgaaatttttttcgttttatttttatttttgttttcatagtgcatgttcat 9896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9841 TCTTACAACACATGAAATTTTTTCGTTTTATTTTATTTTGTTTTCATAGTGCATGTTCAT 9900

Qy    9897 ttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggcagcaaag 9956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9901 TTCTACTCACAAACATGTTCTTGGTGTATTTCTTATGCAAACAATCTTCAGGCAGCAAAG 9960

Qy    9957 atgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||||| |||||||||||||||
Db    9961 ATGTCTGTTACATCTAAACTTGAATAATAAAGTTTTACCACCAGTTACACA 10011


RESULT   2
AC023142/c
LOCUS       AC023142   159505 bp    DNA              HTG       20-AUG-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-558B2, WORKING DRAFT SEQUENCE,
            19 unordered pieces.
ACCESSION   AC023142
VERSION     AC023142.4  GI:9795980
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159505)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159505)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 11, 2000 this sequence version replaced gi:9558667.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0558B02
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 152349 bases at least Q40
            Consensus quality: 154057 bases at least Q30
            Consensus quality: 154743 bases at least Q20
            Insert size: 141000; agarose-fp
            Insert size: 157705; sum-of-contigs
            Quality coverage: 4.80 in Q20 bases; agarose-fp
            Quality coverage: 4.85 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1432: contig of 1432 bp in length
            *     1433     1532: gap of unknown length
            *     1533     2904: contig of 1372 bp in length
            *     2905     3004: gap of unknown length
            *     3005     5152: contig of 2148 bp in length
            *     5153     5252: gap of unknown length
            *     5253     7331: contig of 2079 bp in length
            *     7332     7431: gap of unknown length
            *     7432    11987: contig of 4556 bp in length
            *    11988    12087: gap of unknown length
            *    12088    16583: contig of 4496 bp in length
            *    16584    16683: gap of unknown length
            *    16684    21578: contig of 4895 bp in length
            *    21579    21678: gap of unknown length
            *    21679    26807: contig of 5129 bp in length
            *    26808    26907: gap of unknown length
            *    26908    31857: contig of 4950 bp in length
            *    31858    31957: gap of unknown length
            *    31958    37710: contig of 5753 bp in length
            *    37711    37810: gap of unknown length
            *    37811    43884: contig of 6074 bp in length
            *    43885    43984: gap of unknown length
            *    43985    53419: contig of 9435 bp in length
            *    53420    53519: gap of unknown length
            *    53520    63008: contig of 9489 bp in length
            *    63009    63108: gap of unknown length
            *    63109    75765: contig of 12657 bp in length
            *    75766    75865: gap of unknown length
            *    75866    90579: contig of 14714 bp in length
            *    90580    90679: gap of unknown length
            *    90680   104807: contig of 14128 bp in length
            *   104808   104907: gap of unknown length
            *   104908   118444: contig of 13537 bp in length
            *   118445   118544: gap of unknown length
            *   118545   135079: contig of 16535 bp in length
            *   135080   135179: gap of unknown length
            *   135180   159505: contig of 24326 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-558B2"
     misc_feature    1. .1432
                     /note="assembly_name:Contig10"
     misc_feature    1533. .2904
```

```
                     /note="assembly_name:Contig11"
     misc_feature    3005. .5152
                     /note="assembly_name:Contig12"
     misc_feature    5253. .7331
                     /note="assembly_name:Contig13"
     misc_feature    7432. .11987
                     /note="assembly_name:Contig14"
     misc_feature    12088. .16583
                     /note="assembly_name:Contig15"
     misc_feature    16684. .21578
                     /note="assembly_name:Contig16"
     misc_feature    21679. .26807
                     /note="assembly_name:Contig17"
     misc_feature    26908. .31857
                     /note="assembly_name:Contig18"
     misc_feature    31958. .37710
                     /note="assembly_name:Contig19"
     misc_feature    37811. .43884
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:right"
     misc_feature    43985. .53419
                     /note="assembly_name:Contig21"
     misc_feature    53520. .63008
                     /note="assembly_name:Contig22"
     misc_feature    63109. .75765
                     /note="assembly_name:Contig23"
     misc_feature    75866. .90579
                     /note="assembly_name:Contig24"
     misc_feature    90680. .104807
                     /note="assembly_name:Contig25"
     misc_feature    104908. .118444
                     /note="assembly_name:Contig26"
     misc_feature    118545. .135079
                     /note="assembly_name:Contig27"
     misc_feature    135180. .159505
                     /note="assembly_name:Contig28"
BASE COUNT    50304 a  28279 c  27746 g  51369 t   1807 others
ORIGIN


  Query Match             64.7%;  Score 6476.6;  DB 68;  Length 159505;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 6576;  Conservative    0;  Mismatches    9;  Indels   14;  Gaps    8;

Qy    3421 gactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattattaggaggag 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70101 GACTCCCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAG 70042

Qy    3481 aacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagag 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70041 AACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAG 69982

Qy    3541 gggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatct 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69981 GGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCT 69922

Qy    3601 aactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccaggg 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69921 AACTGAAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGG 69862

Qy    3661 ttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggatat 3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69861 TTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATAT 69802

Qy    3721 tcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagtactggaga 3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69801 TCACAAAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGA 69742

Qy    3781 tccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtcctat 3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69741 TCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTAT 69682

Qy    3841 agaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattga 3900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69681 AGAAAAGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTGA 69622

Qy    3901 aaagtaccagtacccacttttttggacttcccttt-tacataatgacttccagagtgaagc 3959
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   69621 AAAGTACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAGC 69562

Qy    3960 tgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactactt 4019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69561 TGATTGGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACTT 69502

Qy    4020 gagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttcaa 4079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69501 GAGTCACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCAA 69442

Qy    4080 tctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcgat 4139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69441 TCTGCCTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGAT 69382

Qy    4140 caagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcacc 4199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69381 CAAGCATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCACC 69322

Qy    4200 accaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaag 4259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69321 ACCAAATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAAG 69262

Qy    4260 tactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaagt 4319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69261 TACTCAAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAGT 69202

Qy    4320 gatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatatg 4379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69201 GATGTATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATG 69142

Qy    4380 ccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcatag 4439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69141 CCAGCATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATAG 69082

Qy    4440 gaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagcacttagca 4499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69081 GAACAATGCACAAGTGGAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCA 69022

Qy    4500 caattaaatagaaatagggttttcttgatgggaattcaatagcttgtaatgtcttatgaag 4559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69021 CAATTAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAG 68962

Qy    4560 acctattaaaaaaatacttcatagagcctgccttatccaacatgaaattcccttctttg 4619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68961 ACCTATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTTG 68902

Qy    4620 ttattctttctttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatga 4679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68901 TTATTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGA 68842

Qy    4680 gaaagaatggaagatggtaaacaatcacttttttaaaacctgttaagtcaaaaccatcttg 4739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68841 GAAAGAATGGAAGATGGTAAACAATCACTTTTTAAAACCTGTTAAGTCAAAACCATCTTG 68782

Qy    4740 gctaatatgtactggggaaataatccataagagatatcaccagactagaattaatatatt 4799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68781 GCTAATATGTACTGGGGAAATAATCCATAAGAGATATCACCAGACTAGAATTAATATATT 68722

Qy    4800 tataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatactaa 4859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68721 TATAAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGTTTACATATTATTATACTAA 68662

Qy    4860 agcagtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaat 4919
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   68661 AGCAGTACTGGACCGGCCATTGGACCATTTGTTCCAAAACCCATAAATTGTTGCCTAAAT 68602

Qy    4920 ttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaaa 4979
```

```
Db  68601 TTATAATGATCATGAAACCCTAGGCAGAGGAGGAGAAATTGAAGGTCCAGGGCAATGAAA 68542

Qy   4980 gaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagattttgacct 5039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68541 GAAAAATGGCGCCCTCTCAATTTAGTCTTCTCTCATTGGCCATGTTTCAGATTTTGACCT 68482

Qy   5040 agaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtgg 5099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68481 AGAAATGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTGG 68422

Qy   5100 cacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctgatattttt 5159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68421 CACGCTGTTTTTACCCAGCCCTGCCTGTACATACACATGCACACCCTCTCTGATATTTTT 68362

Qy   5160 gtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattcattttgtc 5219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68361 GTCCTTTAGATGTTCAAATACTCAGTAGTCCTTTTGTTTGCGGTTTAGATTCATTTTGTC 68302

Qy   5220 cacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgatttcatttta 5279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68301 CACACATGTACCCATTTTAAAAAACAATGTCCTCGATGCTTCTGTAGTGATTTCATTTTA 68242

Qy   5280 gccaggtatttcttttcttgtgtgtgatgaaccagtatggatttgcttttctaagcctcct 5339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68241 GCCAGGTATTTCTTTCTTGTGTGTGATGAACCAGTATGGATTTGCTTTTCTAAGCCTCCT 68182

Qy   5340 gttggttactaatctcacttggcacattataactaaaggaatcccctcaattcaaaagca 5399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68181 GTTGGTTACTAATCTCACTTGGCACATTATAACTAAAGGAATCCCCTCAATTCAAAAGCA 68122

Qy   5400 tagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttcttc 5459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68121 TAGATGGATACAAATGTCAGACCGTGGGTTTAATTTGTTTAGAACACATGGCATTTCTTC 68062

Qy   5460 acaaggtaacctgctgtatttatttatttttcttttggttaaatataatttccaaactttg 5519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68061 ACAAGGTAACCTGCTGTATTTATTTATTTTCTTTTGGTTAAATATAATTTCCAAACTTTG 68002

Qy   5520 tggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgtaccagcc 5579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  68001 TGGTCAGGCAGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAGCC 67942

Qy   5580 aatcccttcattaaatgtatacagatttagttaagtagcattaaataggattcttagaag 5639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67941 AATCCCTTCATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAG 67882

Qy   5640 tatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaagggaaag 5699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67881 TATGTCCTCATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAG 67822

Qy   5700 ctcctcagcataactgctcagggaaatagggctaaataactgaacattaaataattggtt 5759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67821 CTCCTCAGCATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGGTT 67762

Qy   5760 aaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactgact 5819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67761 AAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGACT 67702

Qy   5820 ttaataatttgcattatattgtcccaaccagtagtttatttttgccacggagatgtaga 5879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67701 TTAATAATTTGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTAGA 67642

Qy   5880 agatattacaagctactggatgcactgtcagattaacttatttcattaaagaagttggga 5939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67641 AGATATTACAAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGA 67582

Qy   5940 gaacaaatagg-aaaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaa 5998
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  67581 GAACAAATAGGAAAAAAAAAACTTATTTTTCTAGTAAATATTAATGTATTACATTTCAAA 67522

Qy   5999 taatggtgcctgacatattgaataattattttctacagtgtacgtatgcaacaaagatat 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  67521 TAATGGTGCCTGACATATTGAATAATTATTTTCTACAGTGTACGTATGCAACAAAGATAT 67462

Qy   6059 tccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaaatgtagc 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67461 TCCATCATGCATTAGAGTCAGTTCTGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAGC 67402

Qy   6119 aaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtg 6178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67401 AAACAAGGTAATGAAGCAACTATTTCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGTG 67342

Qy   6179 tgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatgg 6238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67341 TGCATTAAAGTTGTAAACGGTAACATGAAACAAATGAAAGTTCTTGCTATAATGGTATGG 67282

Qy   6239 aaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaaagggtagcactta 6298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67281 AAAACAAGAAGGAAATGAAAATATTTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTTA 67222

Qy   6299 ttcattccaagtacttttttttttttaatttttaagctcttaactcacattgttatgctt 6358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67221 TTCATTCCAAGTACTTTTTTTTTTTTAATTTTTAAGCTCTTAACTCACATTGTTATGCTT 67162

Qy   6359 aagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaaacatttc 6418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67161 AAGATGATAAACATATATCCTCTTTTTATTGCTTTGTCTATGTTTCATATGAAACATTTC 67102

Qy   6419 agaaattattttgataagtgttgctggaatctgcaacgctgat-ttttttttgcattctg 6477
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  67101 AGAAATTATTTTGATAAGTGTTGCTGGAATCTGCAACGCTGATGTTTTTTTTTGCATTCTG 67042

Qy   6478 tagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattgttttgt 6537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  67041 TAGTCGCATTTGCACTCCATTTTTACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGT 66982

Qy   6538 ttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggt 6597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66981 TTGGGTTTTGTTTCTTTTTCACAGTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGT 66922

Qy   6598 agagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatg 6657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66921 AGAGTTCAACCAGTTCGTGACTGTTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATG 66862

Qy   6658 ttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagag 6717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66861 TTGTGTTGTCATTTTCATTTTTGCATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAG 66802

Qy   6718 aaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaacgggcct 6777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66801 AAAGCAAGAGACAGAAATCAGGACTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCT 66742

Qy   6778 tattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaagaagtgg 6837
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  66741 TATTCTGATCTCACCTGACGCGTAGCTCTAATATTCACATAAACTGAAATAAAGAAGTGG 66682

Qy   6838 aatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaattttgat 6897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66681 AATGAGGAGCTTTGACATTCAAATTATGTGATGTAATTTATCTTCCTTAGGAATTTTGAT 66622

Qy   6898 ggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaaaga 6957
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66621 GGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAAAGA 66562

Qy   6958 gaggagattcccccaaacaacaatatttaatttttcttagtaaaaagaataacagaatgca 7017
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66561 GAGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAATGCA 66502

Qy   7018 tcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacaccaga 7077
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66501 TCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACCAGA 66442

Qy   7078 agtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactcattaag 7137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66441 AGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATTAAG 66382
```

```
Qy   7138 caatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttt 7197
          ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db  66381 CAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTGTGCATGTATGCCTTTTT 66322

Qy   7198 attttcattcaatttttttttttctcagacagacatagtagtatcaactagcattggaa 7257
          |||||||||||||||||||||||||||||||||||||||||| | ||||||||||||||
Db  66321 ATTTTCATTCAATTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTGGAA 66262

Qy   7258 aatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaatattttt 7317
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  66261 AATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAGAATATTTTT 66202

Qy   7318 ggatagcgttgacacgatagatcttattccatacttctttattattgataatttttatttt 7377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66201 GGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTT 66142

Qy   7378 catttttgctttcattattatacatattttggtggagaagaggttgggctttttttgaaa 7437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66141 CATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAA 66082

Qy   7438 gagacaaaaatttattataacactaaacactccttttttgacatattaaagccttttattc 7497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66081 GAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTC 66022

Qy   7498 catctctcaagatatattataaaatttatttttttaatttaagatttctgaattatttta 7557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66021 CATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATTTTA 65962

Qy   7558 tcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaattt 7617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65961 TCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTTAATGAGGAATTT 65902

Qy   7618 catgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtg 7677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65901 CATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTG 65842

Qy   7678 ccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatatt 7737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65841 CCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATT 65782

Qy   7738 ctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattcttcc 7797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65781 CTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCC 65722

Qy   7798 ttttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatagaaa 7857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65721 TTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAA 65662

Qy   7858 aatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatgactc 7917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65661 AATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTC 65602

Qy   7918 caagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg-aaaga 7976
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  65601 CAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAAAGA 65542

Qy   7977 gtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaagatag 8036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65541 GTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGATAG 65482

Qy   8037 tgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagcatg 8096
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65481 TGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATG 65422

Qy   8097 gggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctcttt 8156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65421 GGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTT 65362

Qy   8157 gttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctattatg 8216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65361 GTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATG 65302

Qy   8217 atccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctgagg 8276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65301 ATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGG 65242

Qy   8277 aaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattattagc 8336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65241 AAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGC 65182

Qy   8337 ctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgct 8396
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  65181 CTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACC-GTCTTGAGACCTGTGAACTGCT 65123

Qy   8397 gccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccagtgg 8456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65122 GCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGG 65063

Qy   8457 tggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgcc 8516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65062 TGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCC 65003

Qy   8517 ttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc-aaaaa 8575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  65002 TTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAAAAA 64943

Qy   8576 aaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgctga 8635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64942 AAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGCTGA 64883

Qy   8636 ctgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaaat 8695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64882 CTGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTAAAT 64823

Qy   8696 attctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatgag 8755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64822 ATTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAATGAG 64763

Qy   8756 aagaagatagaaggacgcccgtcgggactttggagggcattgttattttcccaaagaaag 8815
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  64762 AAGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCAGTGTTATTTTCCCAAAGAAAG 64703

Qy   8816 acggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagtact 8875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64702 ACGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACT 64643

Qy   8876 gtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcatcc 8935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64642 GTTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCC 64583

Qy   8936 cagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagtatg 8995
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64582 CAGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATG 64523

Qy   8996 tcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgtac 9055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64522 TCACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATGTAC 64463

Qy   9056 attcatgtgttcttcttgaaaggaaagggagagctgtaagcttcactctgtcctacaccgg 9115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64462 ATTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGG 64403

Qy   9116 agaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaaattgt 9175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64402 AGAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTATCAGAGAAAATTGT 64343

Qy   9176 ccttctagagcatagagtcccaaaactcaattctggttttcccctg-------ttttttt 9228
          |||||||||||||||||||||||||||||||||||||||||||||        |||||||
Db  64342 CCTTCTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCCTGTTTTTTTTTTTTTT 64283

Qy   9229 ttttttttttttttcccaacatatgaactgcagcatatcacttttttcttttttgtgcctcag 9288
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64282 TTTTTTTTTTTTTCCCAACATATGAACTGCAGCATATCACTTTTTCTTTTTGTGCCTCAG 64223

Qy   9289 gttcctcacctgtaaaattgaaaaatatatgtattaataatattattaataataataatg 9348
```

```
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64222 GTTCCTCAGCTGTAAAATTGAAAAATATATGTATTAATAATATTATTAATAATAATAATG 64163

Qy   9349 gtaatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggag 9408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64162 GTAATGTAGTACTTGTTTGTAAAGCACTTTGAGATCCTTGGTTGAAAGGCACCATAGGAG 64103

Qy   9409 tgccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggccagg 9468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64102 TGCCAAGTATTATTATGTGGCCAAGGGGTTATTTAAACTGTCAGTTCCCAAAGGCCAGG 64043

Qy   9469 aaaggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagt 9528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64042 AAAGGTTGGGGTCATTTTTCTTAAAGACGAGCTGTAAATATCAACTAGGCAGCCAATAGT 63983

Qy   9529 gttgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggct 9588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63982 GTTGACTATGAAGATGCAAAACTATTACTAGGCTGATAAAATCATAGTTTCTTAATGGCT 63923

Qy   9589 accaataaggcaaatatcacaataataaacgccaaattccttagggcggactatttgaca 9648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63922 ACCAATAAGGCAAATATCACAATAATAAACGCCAAATTCCTTAGGGCGGACTATTTGACA 63863

Qy   9649 accacatggaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaa 9708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63862 ACCACATGGAAAACTTTGGGGGAGGCATGAGGGGGGAACATCTCAAAATGCCAATGTAAA 63803

Qy   9709 atttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgt 9768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63802 ATTTAACTTACAGCAATATTCACCAGCAGAAAATGTCTTTCATATGGAATGATTTCATGT 63743

Qy   9769 tgctaagaaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataat 9828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63742 TGCTAAGAAAAAGAATTCAATTTGTAGTCCTGATTTGAATACTAGAATGTTGGCTATAAT 63683

Qy   9829 agttctgttcttacaacacatgaaatttttttcgttttatttttatttttgttttcatagtgc 9888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63682 AGTTCTGTTCTTACAACACATGAAATTTTTTCGTTTTATTTTATTTTGTTTTCATAGTGC 63623

Qy   9889 atgttcatttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcagg 9948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  63622 ATGTTCATTTCTACTCACAAACATGTTCTTGGTGTATTTCTTATGCAAACAATCTTCAGG 63563

Qy   9949 cagcaaagatgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  63562 CAGCAAAGATGTCTGTTACATCTAAACTTGAATAATAAAGTTTTACCACCAGTTACACA 63504


RESULT   3
AC013815
LOCUS       AC013815   167123 bp    DNA              HTG       07-JUN-2000
DEFINITION  Homo sapiens clone RP11-21J16, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC013815
VERSION     AC013815.4  GI:8316201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Calagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 7, 2000 this sequence version replaced gi:7341965.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4033
                Center clone name: 21_J_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160651 bases at least Q40
                Consensus quality: 163054 bases at least Q30
                Consensus quality: 164289 bases at least Q20
                Insert size: 167000; agarose-fp
                Insert size: 166123; sum-of-contigs
                Quality coverage: 6.6 in Q20 bases; agarose-fp
                Quality coverage: 6.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1116: contig of 1116 bp in length
            *     1117 1216: gap of      100 bp
            *     1217     2561: contig of 1345 bp in length
            *     2562 2661: gap of      100 bp
            *     2662     3754: contig of 1093 bp in length
            *     3755 3854: gap of      100 bp
            *     3855     4514: contig of 660 bp in length
            *     4515 4614: gap of      100 bp
            *     4615    11771: contig of 7157 bp in length
            *    11772 11871: gap of      100 bp
            *    11872    22787: contig of 10916 bp in length
            *    22788 22887: gap of      100 bp
            *    22888    31860: contig of 8973 bp in length
            *    31861 31960: gap of      100 bp
            *    31961    53736: contig of 21776 bp in length
            *    53737 53836: gap of      100 bp
            *    53837    81814: contig of 27978 bp in length
            *    81815 81914: gap of      100 bp
            *    81915   123726: contig of 41812 bp in length
            *   123727 123826: gap of      100 bp
            *   123827   167123: contig of 43297 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167123
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21J16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1116
                     /note="assembly_fragment"
     misc_feature    1217. .2561
                     /note="assembly_fragment"
     misc_feature    2662. .3754
```

```
                    /note="assembly_fragment"
     misc_feature   3855. .4514
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
     misc_feature   4615. .11771
                    /note="assembly_fragment"
     misc_feature   11872. .22787
                    /note="assembly_fragment"
     misc_feature   22888. .31860
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   31961. .53736
                    /note="assembly_fragment"
     misc_feature   53837. .81814
                    /note="assembly_fragment"
     misc_feature   81915. .123726
                    /note="assembly_fragment"
     misc_feature   123827. .167123
                    /note="assembly_fragment"
BASE COUNT    51442 a  29991 c  31207 g  53480 t   1003 others
ORIGIN
```

Query Match 61.0%; Score 6099.8; DB 63; Length 167123;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 6275; Conservative 0; Mismatches 132; Indels 16; Gaps 8;

```
Qy   3421 gactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattattaggaggag 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25438 GACTCCCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAG 25497

Qy   3481 aacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagag 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25498 AACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAG 25557

Qy   3541 gggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatct 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25558 GGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCT 25617

Qy   3601 aactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccaggg 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25618 AACTGAAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGG 25677

Qy   3661 ttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggatat 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25678 TTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATAT 25737

Qy   3721 tcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagtactggaga 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25738 TCACAAAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGA 25797

Qy   3781 tccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtcctat 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25798 TCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTAT 25857

Qy   3841 agaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattga 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25858 AGAAAAGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTGA 25917

Qy   3901 aaagtaccagtacccactttttggacttcccttt-tacataatgacttccagagtgaagc 3959
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  25918 AAAGTACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAGC 25977

Qy   3960 tgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactactt 4019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25978 TGATTGGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACTT 26037

Qy   4020 gagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttcaa 4079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26038 GAGTCACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCAA 26097

Qy   4080 tctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcgat 4139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26098 TCTGCCTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGAT 26157

Qy   4140 caagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcacc 4199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26158 CAAGCATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCACC 26217

Qy   4200 accaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaag 4259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26218 ACCAAATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAAG 26277

Qy   4260 tactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaagt 4319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26278 TACTCAAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAGT 26337

Qy   4320 gatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatatg 4379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26338 GATGTATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATG 26397

Qy   4380 ccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcatag 4439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26398 CCAGCATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATAG 26457

Qy   4440 gaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaacccttagcacttagca 4499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26458 GAACAATGCACAAGTGGAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCA 26517

Qy   4500 caattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaag 4559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26518 CAATTAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAG 26577

Qy   4560 acctattaaaaaaatacttcatagagcctgccttatccaacatgaaattccccttcttttg 4619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26578 ACCTATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTTG 26637

Qy   4620 ttattctttctttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatga 4679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26638 TTATTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGA 26697

Qy   4680 gaaagaatggaagatggtaaacaatcacttttttaaaaccttgttaagtcaaaaccatcttg 4739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26698 GAAAGAATGGAAGATGGTAAACAATCACTTTTTAAAACCTGTTAAGTCAAAACCATCTTG 26757

Qy   4740 gctaatatgtactggggaaataatccataagagatatcaccagactagaattaatatatt 4799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26758 GCTAATATGTACTGGGGAAATAATCCATAAGAGATATCACCAGACTAGAATTAATATATT 26817

Qy   4800 tataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatactaa 4859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26818 TATAAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGTTTACATATTATTATACTAA 26877

Qy   4860 agcagtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaat 4919
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  26878 AGCAGTACTGGACCGGCCATTGGACCATTTGTTCCAAAACCCATAAATTGTTGCCTAAAT 26937

Qy   4920 ttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaaa 4979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26938 TTATAATGATCATGAAACCCTAGGCAGAGGAGGAGAAATTGAAGGTCCAGGGCAATGAAA 26997

Qy   4980 gaaaaatggcgcccctctcaatttagtcttctctcattggccatgtttcagattttgacct 5039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26998 GAAAAATGGCGCCCTCTCAATTTAGTCTTCTCTCATTGGCCATGTTTCAGATTTTGACCT 27057

Qy   5040 agaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtgg 5099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27058 AGAAATGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTGG 27117

Qy   5100 cacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctgatattttt 5159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27118 CACGCTGTTTTTACCCAGCCCTGCCTGTACATACACATGCACACCCTCTCTGATATTTTT 27177

Qy   5160 gtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattcattttgtc 5219
```

```
Db  27178 GTCCTTTAGATGTTCAAATACTCAGTAGTCCTTTTGTTTGCGGTTTAGATTCATTTTGTC 27237

Qy   5220 cacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgatttcatttta 5279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27238 CACACATGTACCCATTTTAAAAAACAATGTCCTCGATGCTTCTGTAGTGATTTCATTTTA 27297

Qy   5280 gccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaagcctcct 5339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27298 GCCAGGTATTTCTTTCTTGTGTGTGATGAACCAGTATGGATTTGCTTTTCTAAGCCTCCT 27357

Qy   5340 gttggttactaatctcacttggcacattataactaaaggaatcccctcaattcaaaagca 5399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27358 GTTGGTTACTAATCTCACTTGGCACATTATAACTAAAGGAATCCCCTCAATTCAAAAGCA 27417

Qy   5400 tagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttcttc 5459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27418 TAGATGGATACAAATGTCAGACCGTGGGTTTAATTTGTTTAGAACACATGGCATTTCTTC 27477

Qy   5460 acaaggtaacctgctgtatttatttatttttcttttggttaaatataatttccaaacttg 5519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27478 ACAAGGTAACCTGCTGTATTTATTTATTTTCTTTTGGTTAAATATAATTTCCAAACTTTG 27537

Qy   5520 tggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgtaccagcc 5579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27538 TGGTCAGGCAGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAGCC 27597

Qy   5580 aatcccttcattaaatgtatacagatttagttaagtagcattaaataggattcttagaag 5639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27598 AATCCCTTCATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAG 27657

Qy   5640 tatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaagggaaag 5699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27658 TATGTCCTCATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAG 27717

Qy   5700 ctcctcagcataactgctcagggaaatagggctaaataactgaacattaaataattggtt 5759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27718 CTCCTCAGCATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGGTT 27777

Qy   5760 aaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactgact 5819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27778 AAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGACT 27837

Qy   5820 ttaataatttgcattatattgtcccaaccagtagtttatttttgccacggagatgtaga 5879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27838 TTAATAATTTGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTAGA 27897

Qy   5880 agatattacaagctactggatgcactgtcagattaacttatttcattaaagaagttggga 5939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27898 AGATATTACAAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGA 27957

Qy   5940 gaacaaatagg-aaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaa 5998
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  27958 GAACAAATAGGAAAAAAAAACTTATTTTTCTAGTAAATATTAATGTATTACATTTCAAA 28017

Qy   5999 taatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaacaaagatat 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28018 TAATGGTGCCTGACATATTGAATAATTATTTTCTACAGTGTACGTATGCAACAAAGATAT 28077

Qy   6059 tccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaaatgtagc 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28078 TCCATCATGCATTAGAGTCAGTTCTGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAGC 28137

Qy   6119 aaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtg 6178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28138 AAACAAGGTAATGAAGCAACTATTTCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGTG 28197

Qy   6179 tgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatgg 6238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28198 TGCATTAAAGTTGTAAACGGTAACATGAAACAAATGAAAGTTCTTGCTATAATGGTATGG 28257

Qy   6239 aaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcactta 6298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  28258 AAAACAAGAAGGAAATGAAAATATTTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTTA 28317

Qy   6299 ttcattccaagtacttttttttttttaatttttaagctcttaactcacattgttatgctt 6358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28318 TTCATTCCAAGTACTTTTTTTTTTTTAATTTTTAAGCTCTTAACTCACATTGTTATGCTT 28377

Qy   6359 aagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaaacatttc 6418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28378 AAGATGATAAACATATATCCTCTTTTTATTGCTTTGTCTATGTTTCATATGAAACATTTC 28437

Qy   6419 agaaattattttgataagtgttgctggaatctgcaacgctgat-ttttttttgcattctg 6477
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db  28438 AGAAATTATTTTGATAAGTGTTGCTGGAATCTGCAACGCTGATGTTTTTTTTGCATTCTG 28497

Qy   6478 tagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattgttttgt 6537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28498 TAGTCGCATTTGCACTCCATTTTTACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGT 28557

Qy   6538 ttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggt 6597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28558 TTGGGTTTTGTTTCTTTTTCACAGTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGT 28617

Qy   6598 agagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatg 6657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28618 AGAGTTCAACCAGTTCGTGACTGTTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATG 28677

Qy   6658 ttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagag 6717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28678 TTGTGTTGTCATTTTCATTTTTGCATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAG 28737

Qy   6718 aaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaacgggcct 6777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28738 AAAGCAAGAGACAGAAATCAGGACTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCT 28797

Qy   6778 tattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaagaagtgg 6837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28798 TATTCTGATCTCACCTGTCGCGTAGCTCTAATATTCACATAAACTGAAATAAAGAAGTGG 28857

Qy   6838 aatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaattttgat 6897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28858 AATGAGGAGCTTTGACATTCAAATTATGTGATGTAATTTATCTTCCTTAGGAATTTTGAT 28917

Qy   6898 ggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaaaga 6957
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28918 GGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAAAGA 28977

Qy   6958 gaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaatgca 7017
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28978 GAGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAATGCA 29037

Qy   7018 tcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacaccaga 7077
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29038 TCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACCAGA 29097

Qy   7078 agtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactcattaag 7137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29098 AGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATTAAG 29157

Qy   7138 caatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttt 7197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29158 CAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTTTTT 29217

Qy   7198 attttcattcaatttttttttttttctcagacagacatagtagtatcaactagcattggaa 7257
          ||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||||
Db  29218 ATTTTCATTCAATTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTGGAA 29277

Qy   7258 aatacatatcactattcttggaatatttatggtcagtctactttttagtaaaatatttttt 7317
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  29278 AATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAGAATATTTTT 29337

Qy   7318 ggatagcgttgacacgatagatcttattccatacttctttattattgataattttatttt 7377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29338 GGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTT 29397
```

```
Qy   7378 cattttttgctttcattattatacatattttggtggagaagaggttgggctttttttgaaa 7437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29398 CATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAA 29457

Qy   7438 gagacaaaaatttattataacactaaacactccttttttgacatattaaagcctttattc 7497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29458 GAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTC 29517

Qy   7498 catctctcaagatatattataaaatttatttttttaatttaagatttctgaattatttta 7557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29518 CATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATTTTA 29577

Qy   7558 tcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaattt 7617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29578 TCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTTAATGAGGAATTT 29637

Qy   7618 catgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtg 7677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29638 CATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTG 29697

Qy   7678 ccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatatt 7737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29698 CCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATT 29757

Qy   7738 ctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattcttcc 7797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29758 CTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCC 29817

Qy   7798 ttttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatagaaa 7857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29818 TTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAA 29877

Qy   7858 aatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatgactc 7917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29878 AATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTC 29937

Qy   7918 caagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg-aaaga 7976
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  29938 CAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAAAGA 29997

Qy   7977 gtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaagatag 8036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29998 GTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTTAATCAACTTTTAAGATAG 30057

Qy   8037 tgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagcatg 8096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30058 TGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATG 30117

Qy   8097 gggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctcttt 8156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30118 GGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTT 30177

Qy   8157 gttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctattatg 8216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30178 GTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATG 30237

Qy   8217 atccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgcctgagg 8276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30238 ATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGG 30297

Qy   8277 aaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattattagc 8336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30298 AAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGC 30357

Qy   8337 ctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgct 8396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30358 CTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCT 30417

Qy   8397 gccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccagtgg 8456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30418 GCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGG 30477
```

```
Qy   8457 tggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgcc 8516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30478 TGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCC 30537

Qy   8517 ttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc-aaaaa 8575
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  30538 TTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAAAAA 30597

Qy   8576 aaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgctga 8635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30598 AAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGCTGA 30657

Qy   8636 ctgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaaat 8695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30658 CTGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTAAAT 30717

Qy   8696 attctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatgag 8755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30718 ATTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAATGAG 30777

Qy   8756 aagaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaagaaag 8815
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  30778 AAGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCAGTGTTATTTTCCCAAAGAAAG 30837

Qy   8816 acggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagtact 8875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30838 ACGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACT 30897

Qy   8876 gtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcatcc 8935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30898 GTTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCC 30957

Qy   8936 cagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagtatg 8995
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30958 CAGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATG 31017

Qy   8996 tcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgtac 9055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31018 TCACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATGTAC 31077

Qy   9056 attcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacaccgg 9115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31078 ATTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGG 31137

Qy   9116 agaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaaattgt 9175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31138 AGAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTATCAGAGAAAATTGT 31197

Qy   9176 ccttctagagcatagagtcccaaaactcaattctggttttccccctgttttttttttttt 9235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31198 CCTTCTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCCTGTTTTTTTTTTTTTT 31257

Qy   9236 tttttt---------cccaacatatgaactgcagcatatcacttttttcttttttgtgcctc 9286
          ||||||         || | ||||  | | |  | || ||  |||| |||| |  |||
Db  31258 TTTTTTTTTTTTTTCCCAACCATAGAACCGGGGGAATTTCTTTTTTTTTTTTGGGCCCTA 31317

Qy   9287 aggttcctcacctgtaaaattgaaaaatatatgtattaataatattattaataataataa 9346
          |||| ||| |  || ||| | ||||||||| ||| | || ||| || | || || || |
Db  31318 AGGTCCCTAAGTTGAAAATTGGAAAAATATTTGTTTAAAAAATTTTTTAAAAAAAAAAAG 31377

Qy   9347 tggtaatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccatagg 9406
          || || | ||  || ||| | ||||| |||| | | ||| ||| ||| ||| || | ||
Db  31378 GGGAAAGGAAGACCTGGTTGGAAAAGCCCTTTAAAACCCTGGGTGGAAGGGCCCCTTGGG 31437

Qy   9407 agtgccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggcca 9466
          |  | |  | || || || | ||| |||||||| || |||||| || | || || |||||
Db  31438 GGGCCAAGTTTTTTTTTTGGGCCCAGGGGGGTTTTTAAAACTGCCATTCCCAAAGGGCCA 31497

Qy   9467 ggaaaggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaata 9526
          | || ||| |||||| ||||| | ||| || | |||||||| ||| ||||||| ||||||
Db  31498 GAAAGGGTGGGGGTCTTTTTTTTTAAAAAACAACCTGTAAATTTCACCTAGGCACCCAATA 31557

Qy   9527 gtgttgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatgg 9586
```

```
Db  31558 GCGTGGCCTTTGAAAATGCAAAACTTTTCCTAGGCTGAAAAAATCATAGTTTCTTAAGGG 31617

Qy   9587 ctaccaata-aggcaaatatcacaataataaacgccaaattccttagggcggactatttg 9645
          || ||| ||  ||||||| |||||||||||||| |||||||||||||| || || ||||
Db  31618 CTCCCATTAGGGGCAAATTTCACAATAATAAACCCCAAATTCCTTAGGGGGGCCTTTTTG 31677

Qy   9646 acaaccacatggaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgt 9705
           || || | || |||||||||||||||| |||||||||||||||||||| ||||||||
Db  31678 CCACCCCCTTGAAAAACTTTGGGGGAGGCTTGAGGGGGGAACATCTCAAAAGGCCAATGT 31737

Qy   9706 aaaatttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttca 9765
          |||||||||||||||| | ||| ||| || |||||||||||||| ||||| |||| ||
Db  31738 AAAATTTAACTTACAGCATTTTTCCCCAACAAAAAATGTCTTTCATTTGGAAGGATTCCA 31797

Qy   9766 tgttgctaagaaaaagaattcaatttgta-gtcctgatttgaatactagaatgttggcta 9824
          ||| ||||| ||||| |||||||||||| | | |||||||||||||| | || |  |||||
Db  31798 TGTGGCTAAAAAAAAAAATTCAATTTGGAGGCCCTGATTTGAATACCAAAAGGGGGGCTA 31857

Qy   9825 taa 9827
          |||
Db  31858 TAA 31860


RESULT  4
AF264784
LOCUS       AF264784     5507 bp    mRNA            PRI       05-OCT-2000
DEFINITION  Homo sapiens zinc finger protein GC79 mRNA, complete cds.
ACCESSION   AF264784
VERSION     AF264784.1  GI:10644121
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5507)
  AUTHORS   Chang,G.T.G., Steenbeek,M., Schippers,E., Blok,L.J., van
            Weerden,W.M., van Alewijk,D.C., Eussen,B.H., van Steenbrugge,G.J.
            and Brinkmann,A.O.
  TITLE     Characterization of a Zinc-Finger Protein and Its Association With
            Apoptosis in Prostate Cancer Cells
  JOURNAL   J. Natl. Cancer Inst. 92 (17), 1414-1421 (2000)
   PUBMED   10974077
REFERENCE   2  (bases 1 to 5507)
  AUTHORS   Chang,G.T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2000) Endocrinology & Reproduction, Erasmus
            University Rotterdam, P.O.Box 1738, Rotterdam 3000 DR, Netherlands
FEATURES             Location/Qualifiers
     source          1. .5507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23-q24.1"
     CDS             186. .4031
                     /note="prostate cancer gene product"
                     /codon_start=1
                     /product="zinc finger protein GC79"
                     /protein_id="AAG21134.1"
                     /db_xref="GI:10644122"
                     /translation="MVRKKNPPLRNVASEGEGQILEPIGTESKVSGKNKEFSADQMSE
                     NTDQSDAAELNHKEEHSLHVQDPSSSSKKDLKSAVLSEKAGFNYESPSKGGNFPSFPH
                     DEVTDRNMLAFSSPAAGGVCEPLKSPQRAEADDPQDMACTPSGDSLETKEDQKMSPKA
                     TEETGQAQSGQANCQGLSPVSVASKNPQVPSDGGVRLNKSKTDLLVNDNPDPAPLSPE
                     LQDFKCNICGYGYYGNDPTDLIKHFRKYHLGLHNRTRQDAELDSKILALHNMVQFSHS
                     KDFQKVNRSVFSGVLQDINSSRPVLLNGTYDVQVTSGGTFIGIGRKTPDCQGNTKYFR
                     CKFCNFTYMGNSSTELEQHFLQTHPNKIKASLPSSEVAKPSEKNSNKSIPALQSSDSG
                     DLGKWQDKITVKAGDDTPVGYSVPIKPLDSSRQNGTEATSYYWCKFCSFSCESSSSLK
                     LLEHYGKQHGAVQSGGLNPELNDKLSRGSVINQNDLAKSSEGETMTKTDKSSSGAKKK
                     DFSSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLHNIHKCTIKHCPF
                     CPRGLCSPEKHLGEITYPFACRKSNCSHCALLLLHLSPGAAGSSRVKHQCHQCSFTTP
                     DVDVLLFHYESVHESQASDVKQEANHLQGSDGQQSVKESKEHSCTKCDFITQVEEEIS
                     RHYRRAHSCYKCRQCSFTAADTQSLLEHFNTVHCQEQDITTANGEEDGHAISTIKEEP
                     KIDFRVYNLLTPDSKMGEPVSESVVKREKLEEKDGLKEKVWTESSSDDLRNVTWRGAD
                     ILRGSPSYTQASLGLLTPVSGTQEQTKTLRDSPNVEAAHLARPIYGLAVETKGFLQGA
                     PAGGEKSGALPQQYPASGENKSKDESQSLLRRRRGSGVFCANCLTTKTSLWRKNANGG
                     YVCNACGLYQKLHSTPRPLNIIKQNNGEQIIRRRTRKRLNPEALQAEQLNKQQRGSNE
                     EQVNGSPLERRSEDHLTESHQREIPLPSLSKYEAQGSLTKSHSAQQPVLVSQTLDIHK
                     RMQPLHIQIKSPQESTGDPGNSSSVSEGKGSSERGSPIEKYMRPAKHPNYSPPGSPIE
                     KYQYPLFGLPFVHNDFQSEADWLRFWSKYKLSVPGNPHYLSHVPGLPNPCQNYVPYPT
                     FNLPPHFSAVGSDNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVKTEK
                     VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHGDSGPFQCSICQHLCTDKYDFTTHI
                     QRGLHRNNAQVEKNGKPKE"
BASE COUNT     1694 a   1213 c   1234 g   1366 t
ORIGIN


  Query Match             52.7%;  Score 5271;  DB 88;  Length 5507;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 5330;  Conservative    0;  Mismatches   25;  Indels    4;  Gaps    4;

Qy     602 gccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaagaaccccccc 661
           | ||   || |   |  ||   || | |  ||| ||||||||||||||||||||||||||
Db     149 GGCTGCCGAGGGTGTTCTTGACGATTAATCAACAGATATGGTCCGGAAAAAGAACCCCCC 208

Qy     662 tctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggtacagaaag 721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 TCTGAGAAACGTTGCAAGTGAAGGCGAGGGCCAGATCCTGGAGCCTATAGGTACAGAAAG 268

Qy     722 caaggtatctggaaagaacaaagaattctctgcagatcagatgtcagaaaatacggatca 781
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     269 CAAGGTATCTGGAAAGAACAAAGAATTTTCTGCAGATCAGATGTCAGAAAATACGGATCA 328

Qy     782 gagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaagatccatc 841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 GAGTGATGCTGCAGAACTAAATCATAAGGAGGAACATAGCTTGCATGTTCAAGATCCATC 388

Qy     842 ttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggcttcaatta 901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 TTCTAGCAGTAAGAAGGACTTGAAAAGCGCAGTTCTGAGTGAGAAGGCTGGCTTCAATTA 448

Qy     902 tgaaagccccagtaagggaggaaacttttccctcctttccgcatgatgaggtgacagacag 961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 TGAAAGCCCCAGTAAGGGAGGAAACTTTCCCTCCTTTCCGCATGATGAGGTGACAGACAG 508

Qy     962 aaatatgttggctttctcatttccagctgctgggggagtctgtgagcccttgaagtctcc 1021
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     509 AAATATGTTGGCTTTCTCATCTCCAGCTGCTGGGGGAGTCTGTGAGCCCTTGAAGTCTCC 568

Qy    1022 gcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcaggggactcact 1081
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569 GCAAAGAGCAGAGGCAGATGACCCTCAAGATATGGCCTGCACCCCCTCAGGGGACTCACT 628

Qy    1082 ggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggcaagcaca 1141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 GGAGACAAAGGAAGATCAGAAGATGTCACCAAAGGCTACAGAGGAAACAGGGCAAGCACA 688

Qy    1142 gagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaaacccaca 1201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 GAGTGGTCAAGCCAATTGTCAAGGTTTGAGCCCAGTTTCAGTGGCCTCAAAAAACCCACA 748

Qy    1202 agtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactggtgaatga 1261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 AGTGCCTTCAGATGGGGGTGTAAGACTGAATAAATCCAAAACTGACTTACTGGTGAATGA 808

Qy    1262 caacccagacccggcacctctgtctccagagcttcaggactttaaatgcaatatctgtgg 1321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 CAACCCAGACCCGGCACCTCTGTCTCCAGAGCTTCAGGACTTTAAATGCAATATCTGTGG 868

Qy    1322 atatggttactacggcaacgaccccacagatctgattaagcacttccgaaagtatcactt 1381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 ATATGGTTACTACGGCAACGACCCCACAGATCTGATTAAGCACTTCCGAAAGTATCACTT 928

Qy    1382 aggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttggcccttca 1441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 AGGACTGCATAACCGCACCAGGCAAGATGCTGAGCTGGACAGCAAAATCTTGGCCCTTCA 988
```

```
Qy   1442 taacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctgtgttttc 1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    989 TAACATGGTGCAGTTCAGCCATTCCAAAGACTTCCAGAAGGTCAACCGTTCTGTGTTTTC 1048

Qy   1502 tggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacctatgatgt 1561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1049 TGGTGTGCTGCAGGACATCAATTCTTCAAGGCCTGTTTTACTAAATGGGACCTATGATGT 1108

Qy   1562 gcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagattgccaagg 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1109 GCAGGTGACTTCAGGTGGAACATTCATTGGCATTGGACGGAAAACACCAGATTGCCAAGG 1168

Qy   1622 gaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaactcatccac 1681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1169 GAACACCAAGTATTTCCGCTGTAAATTCTGCAATTTCACTTATATGGGCAACTCATCCAC 1228

Qy   1682 cgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctctcccctc 1741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1229 CGAATTAGAACAACATTTTCTTCAGACTCACCCAAACAAAATAAAAGCTTCTCTCCCCTC 1288

Qy   1742 ctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcacttcaatc 1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1289 CTCTGAGGTTGCAAAACCTTCAGAGAAAAACTCTAACAAGTCCATCCCTGCACTTCAATC 1348

Qy   1802 cagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcaggagatga 1861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1349 CAGTGATTCTGGAGACTTGGGAAAATGGCAGGACAAGATAACAGTCAAAGCAGGAGATGA 1408

Qy   1862 cactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaaa-ggtac 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1409 CACTCCTGTTGGGTACTCAGTGCCCATAAAGCCCCTCGATTCCTCTAGACAAAATGGTAC 1468

Qy   1921 agaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcatctagctc 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1469 AGAGGCCACCAGTTACTACTGGTGTAAATTTTGTAGTTTCAGCTGTGAGTCATCTAGCTC 1528

Qy   1981 acttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcggccttaa 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1529 ACTTAAACTGCTAGAACATTATGGCAAGCAGCACGGAGCAGTGCAGTCAGGCGGCCTTAA 1588

Qy   2041 tccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatctagccaa 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1589 TCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCATTAATCAGAATGATCTAGCCAA 1648

Qy   2101 aagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggctaaaaagaa 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1649 AAGTTCAGAAGGAGAGACAATGACCAAGACAGACAAGAGCTCGAGTGGGGCTAAAAAGAA 1708

Qy   2161 ggacttctccagcaagggagccgaggataatatggtaacgagctataattgtcagttctg 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1709 GGACTTCTCCAGCAAGGGAGCCGAGGATAATATGGTAACGAGCTATAATTGTCAGTTCTG 1768

Qy   2221 tgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccacttctccg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1769 TGACTTCCGATATTCCAAAAGCCATGGCCCTGATGTAATTGTAGTGGGGCCACTTCTCCG 1828

Qy   2281 tcattatcaacagctccataacattcacaagtgtaccattaaacactgtccattctgtcc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1829 TCATTATCAACAGCTCCATAACATTCACAAGTGTACCATTAAACACTGTCCATTCTGTCC 1888

Qy   2341 cagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttgcttgtag 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1889 CAGAGGACTTTGCAGCCCAGAAAAGCACCTTGGAGAAATTACTTATCCGTTTGCTTGTAG 1948

Qy   2401 aaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctggggcggctgg 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1949 AAAAAGTAATTGTTCCCACTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGCGGCTGG 2008

Qy   2461 aagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacgtagatgt 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2009 AAGCTCGCGAGTCAAACATCAGTGCCATCAGTGTTCATTCACCACCCCTGACGTAGATGT 2068

Qy   2521 actcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaacaagaagc 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2069 ACTCCTCTTTCACTATGAAAGTGTGCATGAGTCCCAAGCATCGGATGTCAAACAAGAAGC 2128

Qy   2581 aaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaacactcatg 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2129 AAATCACCTGCAAGGATCGGATGGGCAGCAGTCTGTCAAGGAAAGCAAAGAACACTCATG 2188

Qy   2641 taccaaatgtgattttattacccaagtggaagaagagatttcccgacactacaggagagc 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2189 TACCAAATGTGATTTTATTACCCAAGTGGAAGAAGAGATTTCCCGACACTACAGGAGAGC 2248

Qy   2701 acacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagtcactact 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2249 ACACAGCTGCTACAAATGCCGTCAGTGCAGTTTTACAGCTGCCGATACTCAGTCACTACT 2308

Qy   2761 ggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacggcgaaga 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2309 GGAGCACTTCAACACTGTTCACTGCCAGGAACAGGACATCACTACAGCCAACGGCGAAGA 2368

Qy   2821 ggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcagggtctacaa 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2369 GGACGGTCATGCCATATCCACCATCAAAGAGGAGCCCAAAATTGACTTCAGGGTCTACAA 2428

Qy   2881 tctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtgaagagaga 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2429 TCTGCTAACTCCAGACTCTAAAATGGGAGAGCCAGTTTCTGAGAGTGTGGTGAAGAGAGA 2488

Qy   2941 gaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttccagtgatga 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2489 GAAGCTGGAAGAGAAGGACGGGCTCAAAGAGAAAGTTTGGACCGAGAGTTCCAGTGATGA 2548

Qy   3001 ccttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcatacaccca 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2549 CCTTCGCAATGTGACTTGGAGAGGGGCAGACATCCTGCGGGGGAGTCCGTCATACACCCA 2608

Qy   3061 agcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaagactctaag 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2609 AGCAAGCCTGGGGCTGCTGACGCCTGTGTCTGGCACCCAAGAGCAGACAAAGACTCTAAG 2668

Qy   3121 ggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttggctgtgga 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2669 GGATAGTCCCAATGTGGAGGCCGCCCATCTGGCGCGACCTATTTATGGCTTGGCTGTGGA 2728

Qy   3181 aaccaagggattcctgcagggggcgccagctggcggagagaagtctggggccctccccca 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2729 AACCAAGGGATTCCTGCAGGGGGCGCCAGCTGGCGGAGAGAAGTCTGGGGCCCTCCCCCA 2788

Qy   3241 gcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgttacggag 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2789 GCAGTATCCTGCATCGGGAGAAAACAAGTCCAAGGATGAATCCCAGTCCCTGTTACGGAG 2848

Qy   3301 gcgtagaggctccggtgtttttttgtgccaattgcctgaccacaaagacctctctctggcg 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2849 GCGTAGAGGCTCCGGTGTTTTTTGTGCCAATTGCCTGACCACAAAGACCTCTCTCTGGCG 2908

Qy   3361 aaagaatgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagcttcactc 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2909 AAAGAATGCAAATGGCGGATATGTATGCAACGCGTGTGGCCTCTACCAGAAGCTTCACTC 2968

Qy   3421 gactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattattaggaggag 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2969 GACTCCCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAG 3028

Qy   3481 aacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagag 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3029 AACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAG 3088

Qy   3541 gggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatct 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3089 GGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCT 3148

Qy   3601 aactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccaggg 3660
```

```
Db   3149 AACTGAAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGG 3208

Qy   3661 ttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggatat 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3209 TTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATAT 3268

Qy   3721 tcacaaaaggatgcaaccttgcacattcagataaaaagtcctcaggaaagtactggaga 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3269 TCACAAAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGA 3328

Qy   3781 tccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtcctat 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3329 TCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTAT 3388

Qy   3841 agaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattga 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3389 AGAAAAGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTGA 3448

Qy   3901 aaagtaccagtacccactttttggacttcccttt-tacataatgacttccagagtgaagc 3959
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   3449 AAAGTACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAGC 3508

Qy   3960 tgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactactt 4019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3509 TGATTGGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACTT 3568

Qy   4020 gagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttcaa 4079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3569 GAGTCACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCAA 3628

Qy   4080 tctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcgat 4139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3629 TCTGCCTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGAT 3688

Qy   4140 caagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcacc 4199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3689 CAAGCATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCACC 3748

Qy   4200 accaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaag 4259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3749 ACCAAATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAAG 3808

Qy   4260 tactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaagt 4319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3809 TACTCAAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAGT 3868

Qy   4320 gatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatatg 4379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3869 GATGTATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATG 3928

Qy   4380 ccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcatag 4439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3929 CCAGCATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATAG 3988

Qy   4440 gaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagcacttagca 4499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3989 GAACAATGCACAAGTGGAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCA 4048

Qy   4500 caattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaag 4559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4049 CAATTAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAG 4108

Qy   4560 acctattaaaaaaatacttcatagagcctgccttatccaacatgaaattcccttcttttg 4619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4109 ACCTATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTTG 4168

Qy   4620 ttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatga 4679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4169 TTATTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGA 4228

Qy   4680 gaaagaatggaagatggtaaacaatcacttttaaaacctgttaagtcaaaaccatcttg 4739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   4229 GAAAGAATGGAAGATGGTAAACAATCACTTTTTAAAACCTGTTAAGTCAAAACCATCTTG 4288

Qy   4740 gctaatatgtactggggaaataatccataagagatatcaccagactagaattaatatatt 4799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4289 GCTAATATGTACTGGGGAAATAATCCATAAGAGATATCACCAGACTAGAATTAATATATT 4348

Qy   4800 tataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatactaa 4859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4349 TATAAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGTTTACATATTATTATACTAA 4408

Qy   4860 agcagtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaat 4919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 AGCAGTACTGGACTGGCCATTGGACCATTTGTTCCAAAACCCATAAATTGTTGCCTAAAT 4468

Qy   4920 ttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaaa 4979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4469 TTATAATGATCATGAAACCCTAGGCAGAGGAGGAGAAATTGAAGGTCCAGGGCAATGAAA 4528

Qy   4980 gaaaaatggcgcccctctcaatttagtcttctctcattggccatgtttcagattttgacct 5039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4529 GAAAAATGGCGCCCTCTCAATTTAGTCTTCTCTCATTGGCCATGTTTCAGATTTTGACCT 4588

Qy   5040 agaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtgg 5099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 AGAAATGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTGG 4648

Qy   5100 cacgctgtttttacccagccctgcctgtacatacacatgcacaccctctct-gatatttt 5158
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   4649 CACGCTGTTTTTACCCAGCCCTGCCTGTACATACACATGCACACCCTCTCTGGATATTTT 4708

Qy   5159 tgtcctttagatgttcaaatactcagtagtccttttg-tttgcggtttagattcattttg 5217
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   4709 TGTCCTTTAGATGTTCAAATACTCAGTAGTCCTTTTGTTTTGCGGTTTAGATTCATTTTG 4768

Qy   5218 tccacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgatttcattt 5277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 TCCACACATGTACCCATTTTAAAAAACAATGTCCTCGATGCTTCTGTAGTGATTTCATTT 4828

Qy   5278 tagccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaagcctc 5337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 TAGCCAGGTATTTCTTTCTTGTGTGTGATGAACCAGTATGGATTTGCTTTTCTAAGCCTC 4888

Qy   5338 ctgttggttactaatctcacttggcacattataactaaaggaatcccctcaattcaaaag 5397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 CTGTTGGTTACTAATCTCACTTGGCACATTATAACTAAAGGAATCCCCTCAATTCAAAAG 4948

Qy   5398 catagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttct 5457
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   4949 CATAGATGGATACCAATGTCAGACCGTGGGTTTAATTTGTTTAGAACACATGGCATTTCT 5008

Qy   5458 tcacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttccaaactt 5517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5009 TCACAAGGTAACCTGCTGTATTTATTTATTTTCTTTTGGTTAAATATAATTTCCAAACTT 5068

Qy   5518 tgtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgtaccag 5577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5069 TGTGGTCAGGCAGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAG 5128

Qy   5578 ccaatcccttcattaaatgtatacagatttagttaagtagcattaaataggattcttaga 5637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5129 CCAATCCCTTCATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGA 5188

Qy   5638 agtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaagggaa 5697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5189 AGTATGTCCTCATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAA 5248

Qy   5698 agctcctcagcataactgctcagggaaatagggctaaataactgaacattaaataattgg 5757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5249 AGCTCCTCAGCATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGG 5308

Qy   5758 ttaaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactga 5817
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5309 TTAAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGA 5368
```

```
Qy   5818 ctttaataatttgcattatattgtcccaaccagtagtttatttttttgccacggagatgta 5877
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5369 CTTTAATAATTTGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTA 5428

Qy   5878 gaagatattacaagctactggatgcactgtcagattaacttatttcattaaagaagttgg 5937
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5429 GAAGATATTACAAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGG 5488

Qy   5938 gagaacaaataggaaaaaa 5956
          ||||||||| | ||||||
Db   5489 GAGAACAAAAAAAAAAAAA 5507
```

RESULT 5
AK000948
LOCUS AK000948 2596 bp mRNA PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ10086 fis, clone HEMBA1002177, weakly similar to TRANSCRIPTION FACTOR GATA-4.
ACCESSION AK000948
VERSION AK000948.1 GI:7021932
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to mRNA, clone_lib:HEMBA1 clone:HEMBA1002177.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2596)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES Location/Qualifiers
source 1. .2596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1002177"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/tissue_type="whole embryo, mainly head"
/note="cloning vector: pME18SFL3"
CDS 111. .2468
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91441.1"
/db_xref="GI:7021933"
/translation="MTKTDKSSSGAKKKDFSSKGAEDNMVTSYNCQFCDFRYSKSHGP
DVIVVGPLLRHYQQLHNIHKCTIKHCPFCPRGLCSPEKHLGEITYPFACRKSNCSHCA
LLLLHLSPGAAGSSRVKHQCHQCSFTTPDVDVLLFHYESVHESQASDVKQEANHLQGS
DGQQSVKESKEHSCTKCDFITQVEEEISRHYRRAHSCYKCRQCSFTAADTQSLLEHFN
TVHCQEQDITTANGEEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSESVVKREKL
EEKDGLKEKVWTESSSDDLRNVTWRGADILRGSPSYTQASLGLLTPVSGTQEQTKTLR
DSPNVEAAHLARPIYGLAVETKGFLQGAPAGGEKSGALPQQYPASGENKSKDESQSLL
RRRRGSGVFCANCLTTKTSLWRKNANGGYVCNACGLYQKLHSTPRPLNIIKQNNGEQI
IRRRTRKRLNPEALQAEQLNKQQRGSNEEQVNGSPLERRSEDHLTESHQREIPLPSLS
KYEAQGSLTKSHSAQQPVLVSQTLDIHKRMQPLHIQIKSPQESTGDPGNSSSVSEGKG
SSERGSPIEKYMRPAKHPNYSPPGSPIEKYQYPLFGLPFVHNDFQSEADWLRFWSKYK
LSVPGNPHYLSHVPGLPNPCQNYVPYPTFNLPPHFSAVGSDNDIPLDLAIKHSRPGPT
ANGASKEKTKAPPNVKNEGPLNVVKTEKVDRSTQDELSTKCVHCGIVFLDEVMYALHM
SCHGDSGPFQCSICQHLCTDKYDFTTHIQRGLHRNNAQVEKKWKT"
BASE COUNT 805 a 609 c 615 g 567 t
ORIGIN

Query Match 24.7%; Score 2474.4; DB 88; Length 2596;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2497; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
Qy   2106 cagaaggagagacaatgaccaagacagacaagagctcgagtggggctaaaaagaaggact 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 CAGAAGGAGAGACAATGACCAAGACAGACAAGAGCTCGAGTGGGGCTAAAAAGAAGGACT 156

Qy   2166 tctccagcaagggagccgaggataatatggtaacgagctataattgtcagttctgtgact 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 TCTCCAGCAAGGGAGCCGAGGATAATATGGTAACGAGCTATAATTGTCAGTTCTGTGACT 216

Qy   2226 tccgatattccaaaagccatggccctgatgtaattgtagtggggccacttctccgtcatt 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TCCGATATTCCAAAAGCCATGGCCCTGATGTAATTGTAGTGGGGCCACTTCTCCGTCATT 276

Qy   2286 atcaacagctccataacattcacaagtgtaccattaaacactgtccattctgtcccagag 2345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 ATCAACAGCTCCATAACATTCACAAGTGTACCATTAAACACTGTCCATTCTGTCCCAGAG 336

Qy   2346 gactttgcagcccagaaaagcaccttggagaaattacttatccgtttgcttgtagaaaaa 2405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 GACTTTGCAGCCCAGAAAAGCACCTTGGAGAAATTACTTATCCGTTTGCTTGTAGAAAAA 396

Qy   2406 gtaattgttcccactgtgcactcttgcttctgcacttgtctcctggggcggctggaagct 2465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 GTAATTGTTCCCACTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGCGGCTGGAAGCT 456

Qy   2466 cgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacgtagatgtactcc 2525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CGCGAGTCAAACATCAGTGCCATCAGTGTTCATTCACCACCCCTGACGTAGATGTACTCC 516

Qy   2526 tctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaacaagaagcaaatc 2585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 TCTTTCACTATGAAAGTGTGCATGAGTCCCAAGCATCGGATGTCAAACAAGAAGCAAATC 576

Qy   2586 acctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaacactcatgtacca 2645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 ACCTGCAAGGATCGGATGGGCAGCAGTCTGTCAAGGAAAGCAAAGAACACTCATGTACCA 636

Qy   2646 aatgtgattttattacccaagtggaagaagagatttcccgacactacaggagagcacaca 2705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 AATGTGATTTTATTACCCAAGTGGAAGAAGAGATTTCCCGACACTACAGGAGAGCACACA 696

Qy   2706 gctgctacaaatgccgtcagtgcagttttacagctgccgatactcagtcactactggagc 2765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    697 GCTGCTACAAATGCCGTCAGTGCAGTTTTACAGCTGCCGATACTCAGTCACTACTGGAGC 756

Qy   2766 acttcaacactgttcactgccaggaacaggacatcactacagccaacggcgaagaggacg 2825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    757 ACTTCAACACTGTTCACTGCCAGGAACAGGACATCACTACAGCCAACGGCGAAGAGGACG 816

Qy   2826 gtcatgccatatccaccatcaaagaggagcccaaaattgacttcagggtctacaatctgc 2885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    817 GTCATGCCATATCCACCATCAAAGAGGAGCCCAAAATTGACTTCAGGGTCTACAATCTGC 876

Qy   2886 taactccagactctaaaatgggagagccagtttctgagagtgtggtgaagagagagaagc 2945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    877 TAACTCCAGACTCTAAAATGGGAGAGCCAGTTTCTGAGAGTGTGGTGAAGAGAGAGAAGC 936

Qy   2946 tggaagagaaggacgggctcaaagagaaagtttggaccgagagttccagtgatgaccttc 3005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 TGGAAGAGAAGGACGGGCTCAAAGAGAAAGTTTGGACCGAGAGTTCCAGTGATGACCTTC 996

Qy   3006 gcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcatacacccaagcaa 3065
```

```
Db    997 GCAATGTGACTTGGAGAGGGGCAGACATCCTGCGGGGGAGTCCGTCATACACCCAAGCAA 1056

Qy   3066 gcctggggctgctgacgcctgtgtctggcacccaagagcagacaaagactctaagggata 3125
Db   1057 GCCTGGGGCTGCTGACGCCTGTGTCTGGCACCCAAGAGCAGACAAAGACTCTAAGGGATA 1116

Qy   3126 gtcccaatgtggaggccgcccatctggcgcgacctatttatggcttggctgtggaaacca 3185
Db   1117 GTCCCAATGTGGAGGCCGCCCATCTGGCGCGACCTATTTATGGCTTGGCTGTGGAAACCA 1176

Qy   3186 agggattcctgcaggggcgccagctggcggagagaagtctggggccctcccccagcagt 3245
Db   1177 AGGGATTCCTGCAGGGGCGCCAGCTGGCGGAGAGAAGTCTGGGGCCCTCCCCCAGCAGT 1236

Qy   3246 atcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgttacggaggcgta 3305
Db   1237 ATCCTGCATCGGGAGAAAACAAGTCCAAGGATGAATCCCAGTCCCTGTTACGGAGGCGTA 1296

Qy   3306 gaggctccggtgtttttttgtgccaattgcctgaccacaaagacctctctctggcgaaaga 3365
Db   1297 GAGGCTCCGGTGTTTTTTGTGCCAATTGCCTGACCACAAAGACCTCTCTCTGGCGAAAGA 1356

Qy   3366 atgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagcttcactcgactc 3425
Db   1357 ATGCAAATGGCGGATATGTATGCAACGCGTGTGGCCTCTACCAGAAGCTTCACTCGACTC 1416

Qy   3426 ccaggcctttaaacatcattaaacaaaacaacggtgagcagattattaggaggagaacaa 3485
Db   1417 CCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAGAACAA 1476

Qy   3486 gaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagaggggca 3545
Db   1477 GAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAGGGGCA 1536

Qy   3546 gcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatctaactg 3605
Db   1537 GCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCTAACTG 1596

Qy   3606 aaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccagggttcat 3665
Db   1597 AAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGGTTCAT 1656

Qy   3666 tgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggatattcaca 3725
Db   1657 TGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATATTCACA 1716

Qy   3726 aaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagtactggagatccag 3785
Db   1717 AAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGATCCAG 1776

Qy   3786 gaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtcctatagaaa 3845
Db   1777 GAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCCGAGAGAGGCAGTCCTATAGAAA 1836

Qy   3846 agtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattgaaaagt 3905
Db   1837 AGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTGAAAAGT 1896

Qy   3906 accagtacccactttttggacttcccttt-tacataatgacttccagagtgaagctgatt 3964
Db   1897 ACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAGCTGATT 1956

Qy   3965 ggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactacttgagtc 4024
Db   1957 GGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACTTGAGTC 2016

Qy   4025 acgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttcaatctgc 4084
Db   2017 ACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCAATCTGC 2076

Qy   4085 ctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcgatcaagc 4144
Db   2077 CTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGATCAAGC 2136

Qy   4145 attccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcaccaccaa 4204
Db   2137 ATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCACCACCAA 2196

Qy   4205 atgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaagtactc 4264
Db   2197 ATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAAGTACTC 2256

Qy   4265 aagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaagtgatgt 4324
Db   2257 AAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAGTGATGT 2316

Qy   4325 atgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatatgccagc 4384
Db   2317 ATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGC 2376

Qy   4385 atctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcataggaaca 4444
Db   2377 ATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATAGGAACA 2436

Qy   4445 atgcacaagtgg-aaaaaaatggaaaacctaaagagtaaaaccttagcacttagcacaat 4503
Db   2437 ATGCACAAGTGGAAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCACAAT 2496

Qy   4504 taaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaagacct 4563
Db   2497 TAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAGACCT 2556

Qy   4564 attaaaaaaatacttcatagagcctgccttatccaacatg 4603
Db   2557 ATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATG 2596
```

```
RESULT   6
AK021511
LOCUS       AK021511     1587 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ11449 fis, clone HEMBA1001411.
ACCESSION   AK021511
VERSION     AK021511.1  GI:10432707
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
            mRNA, clone_lib:HEMBA1 clone:HEMBA1001411.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1587)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1587
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1001411"
                     /clone_lib="HEMBA1"
                     /dev_stage="embryo, 10 weeks"
                     /tissue_type="whole embryo, mainly head"
                     /note="cloning vector: pME18SFL3"
BASE COUNT       486 a    244 c    276 g    581 t
ORIGIN
```

Query Match 15.7%; Score 1568.6; DB 89; Length 1587;
Best Local Similarity 99.7%; Pred. No. 1.1e-267;
Matches 1582; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy   6985 taattttcttagtaaaaagaataacagaatgcatcgtggcaatccttaagcaacattatc 7044
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TAATTTTCTTAGTAAAAAGAATAACAGAATGCATCGTGGCAATCCTTAAGCAACATTATC 60

Qy   7045 tatgtggactgcttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgaca 7104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TATGTGGACTGCTTAAATCAGCAAAACACCAGAAGTTTGGTTAACTTGGGCAATATGACA 120

Qy   7105 agtattacttttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaa 7164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGTATTACTTTTTGGGCAAAACTACTCATTAAGCAATTTCTCTAGTGTGTCGGACACAAA 180

Qy   7165 taggttctttatttttggcatgtatgccttttttatttttcattcaatttttttttttttctc 7224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TAGGTTCTTTATTTTTGGCATGTATGCCTTTTTTATTTTCATTCAATTTTTTTTTTTTTCTC 240

Qy   7225 agacagacatagtagtatcaactagcattggaaaatacatatcactattcttggaatatt 7284
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
Db    241 AGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATATCACTATTCTTGGAATATT 300

Qy   7285 tatggtcagtctacttttttagtaaaatatttttggatagcgttgacacgatagatcttat 7344
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    301 TATGGTCAGTCTACTTTTTAGTAGAATATTTTTGGATAGCGTTGACACGATAGATCTTAT 360

Qy   7345 tccatacttctttattattgataatttttattttcatttttttgctttcattattatacata 7404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTGCTTTCATTATTATACATA 420

Qy   7405 ttttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaa 7464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAAATTTATTATAACACTAAA 480

Qy   7465 cactccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttt 7524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCAAGATATATTATAAAATTT 540

Qy   7525 atttttttaatttaagatttctgaattatttttatcttaaattgtgattttaaacgagcta 7584
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    541 ATTTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATTGTGATTTTAAGCGAGCTA 600

Qy   7585 ttatggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttctctctt 7644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTATGGTACGGAACTTTTTTTTAATGAGGAATTTCATGATGATTTAGGAATTTTCTCTCTT 660

Qy   7645 ggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccatttaaaa 7704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGAAAAGGCTTCCCCTGTGATGAAAATGATGTGCCAGCTAAAATTGTGTGCCATTTAAAA 720

Qy   7705 actgaaaatattttaaaattatttgtctatattctaaattgagctttggatcaaacttta 7764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTGAAAATATTTTAAAATTATTTGTCTATATTCTAAATTGAGCTTTGGATCAAACTTTA 780

Qy   7765 ggccaggaccagctcatgcgttctcattcttcctttttctcactctttctctcatcactca 7824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGCCAGGACCAGCTCATGCGTTCTCATTCTTCCTTTTCTCACTCTTTCTCTCATCACTCA 840

Qy   7825 cctctgtattcattctgttgtttgggatagaaaaatcataaagagccaacccatctcaga 7884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    841 CCTCTGTATTCATTCTGTTGTTTGGGATAGAAAAATCATAAAGAGCCAACCCATCTCAGA 900

Qy   7885 acgttgtggattgagagagacactacatgactccaagtatatgagaaaaggacagagctc 7944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACGTTGTGGATTGAGAGAGACACTACATGACTCCAAGTATATGAGAAAAGGACAGAGCTC 960

Qy   7945 taattgataactctgtagttcaaaagg-aaagagtatgcccaattctctctacatgacat 8003
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    961 TAATTGATAACTCTGTAGTTCAAAAGGAAAAGAGTATGCCCAATTCTCTCTACATGACAT 1020

Qy   8004 attgagatttttttttaatcaacttttaagatagtgatgttctgttctaaactgttctgtt 8063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ATTGAGATTTTTTTTTAATCAACTTTTAAGATAGTGATGTTCTGTTCTAAACTGTTCTGTT 1080

Qy   8064 ttagtgaaggtagatttttataaaacaagcatggggattcttttctaaggtaatattaat 8123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTAGTGAAGGTAGATTTTTATAAAACAAGCATGGGGATTCTTTTCTAAGGTAATATTAAT 1140

Qy   8124 gagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcacattatgt 8183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GAGAAGGGAAAAAAGTATCTTTAACAGCTCTTTGTTGAAGCCTGTGGTAGCACATTATGT 1200

Qy   8184 ttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctccaaaaat 8243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TTATAATTGCACATGTGCACATAATCTATTATGATCCAATGCAAATACAGCTCCAAAAAT 1260

Qy   8244 attaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataaactgaag 8303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACATTTTTCTTAATAAACTGAAG 1320

Qy   8304 agtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaaaacatca 8363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTTGTGTGGCTGCAAAACATCA 1380

Qy   8364 caaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaattaatgc 8423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTTTAGTAAATAAAATTAATGC 1440

Qy   8424 atttctagaggggggaatatctgccatccagtggtggaaatgtggagtaaagaagctggtg 8483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTG 1500

Qy   8484 gtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaactttagc 8543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGC 1560

Qy   8544 tactgtctttggtttgagagccatggc 8570
          |||||||||||||||||||||||||||
Db   1561 TACTGTCTTTGGTTTGAGAGCCATGGC 1587
```

```
RESULT   7
AK021470
LOCUS       AK021470     1551 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ11408 fis, clone HEMBA1000822.
ACCESSION   AK021470
VERSION     AK021470.1  GI:10432662
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
            mRNA, clone_lib:HEMBA1 clone:HEMBA1000822.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
```

```
REFERENCE   2  (bases 1 to 1551)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1000822"
                     /clone_lib="HEMBA1"
                     /dev_stage="embryo, 10 weeks"
                     /tissue_type="whole embryo, mainly head"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      476 a    238 c    285 g    552 t
ORIGIN
```

Query Match 15.2%; Score 1525; DB 89; Length 1551;
Best Local Similarity 99.8%; Pred. No. 5.9e-260;
Matches 1548; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

```
Qy   5304 gatgaaccagtatggatttgctttttctaagcctcctgttggttactaatctcacttggca 5363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATGAACCAGTATGGATTTGCTTTTCTAAGCCTCCTGTTGGTTACTAATCTCACTTGGCA 60

Qy   5364 cattataactaaaggaatcccctcaattcaaaagcatagatggatacaaatgtcagaccg 5423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CATTATAACTAAAGGAATCCCCTCAATTCAAAAGCATAGATGGATACAAATGTCAGACCG 120

Qy   5424 tgggtttaatttgtttagaacacatggcatttcttcacaaggtaacctgctgtatttatt 5483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGTTTAATTTGTTTAGAACACATGGCATTTCTTCACAAGGTAACCTGCTGTATTTATT 180

Qy   5484 tattttcttttggttaaatataatttccaaacttttgtggtcaggcagcgtctaaggttac 5543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATTTTCTTTTGGTTAAATATAATTTCCAAACTTTGTGGTCAGGCAGCGTCTAAGGTTAC 240

Qy   5544 gttaccacagactgacagttggtatatgtaccagccaatcccttcattaaatgtatacag 5603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTACCACAGACTGACAGTTGGTATATGTACCAGCCAATCCCTTCATTAAATGTATACAG 300

Qy   5604 atttagttaagtagcattaaataggattcttagaagtatgtcctcatagaacttttaata 5663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAGTATGTCCTCATAGAACTTTTAATA 360

Qy   5664 cttaaggctttgtaaaaactatccatgaagggaaagctcctcagcataactgctcaggga 5723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAGCTCCTCAGCATAACTGCTCAGGGA 420

Qy   5724 aatagggctaaataactgaacattaaataattggttaaaggtgctgttagtcgagcctca 5783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATAGGGCTAAATAACTGAACATTAAATAATTGGTTAAAGGTGCTGTTAGTCGAGCCTCA 480

Qy   5784 atgcttgctacaaggatgtatgtacaaggactgactttaataatttgcattatattgtcc 5843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGCTTGCTACAAGGATGTATGTACAAGGACTGACTTTAATAATTTGCATTATATTGTCC 540

Qy   5844 caaccagtagtttatttttgccacggagatgtagaagatattacaagctactggatgca 5903
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAACCAGTAGTTTATTTTTTGCCACGGAGATGTAGAAGATATTACAAGCTACTGGATGCA 600

Qy   5904 ctgtcagattaacttatttcattaaagaagttgggagaacaaatagg-aaaaaaaaactt 5962
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    601 CTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGAGAACAAATAGGAAAAAAAAAACTT 660

Qy   5963 atttttctagtaaatattaatgtattacatttcaaataatggtgcctgacatattgaata 6022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATTTTTCTAGTAAATATTAATGTATTACATTTCAAATAATGGTGCCTGACATATTGAATA 720

Qy   6023 attattttctacagtgtacgtatgcaacaaagatattccatcatgcattagagtcagttc 6082
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATTATTTTCTACAGTGTACGTATGCAACAAAGATATTCCATCATGCATTAGAGTCAGTTC 780

Qy   6083 tggctctgcctagctgtttacatttgcaaatgtagcaaacaaggtaatgaagcaactatt 6142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAGCAAACAAGGTAATGAAGCAACTATT 840

Qy   6143 tctattgcagtagatatccttttgtgtgtgtgtgtgcattaaagttgtaaacggtaac 6202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGTGCATTAAAGTTGTAAACGGTAAC 900

Qy   6203 atgaaacaaatgaaagttcttgctataatggtatggaaaacaagaaggaaatgaaaatat 6262
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATGAAACAAATGAAAGTTCTTGCTATAATGGTATGGAAAACAAGAAGGAAATGAAAATAT 960

Qy   6263 ttttatgcctacttaggaaaaaaagggtagcacttattcattccaagtacttttttttttt 6322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTTATTCATTCCAAGTACTTTTTTTTTT 1020

Qy   6323 ttaatttttaagctcttaactcacattgttatgcttaagatgataaacatatatcctctt 6382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTAATTTTTAAGCTCTTAACTCACATTGTTATGCTTAAGATGATAAACATATATCCTCTT 1080

Qy   6383 tttattgctttgtctatgtttcatatgaaacatttcagaaattattttgataagtgttgc 6442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTATTGCTTTGTCTATGTTTCATATGAAACATTTCAGAAATTATTTTGATAAGTGTTGC 1140

Qy   6443 tggaatctgcaacgctga--ttttttttgcattctgtagtcgcatttgcactccatttt 6500
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGAATCTGCAACGCTGATGTTTTTTTTTGCATTCTGTAGTCGCATTTGCACTCCATTTT 1200

Qy   6501 tacattaattcgcagttgctttgtatcattgttttgtttgggttttgtttcttttttcaca 6560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGTTTGGGTTTTGTTTCTTTTTCACA 1260

Qy   6561 gtgccgggtcttcgtttcttaaagttggatggcaggtagagttcaaccagttcgtgactg 6620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGTAGAGTTCAACCAGTTCGTGACTG 1320

Qy   6621 ttgtagcgaatgaagttaaaaaaatgtctttctgatgttgtgttgtcattttcatttttg 6680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATGTTGTGTTGTCATTTTCATTTTTG 1380

Qy   6681 catttttttgtttgcatattaaaaaaagagaaaagagaaagcaagagacagaaatcagga 6740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAGAAAGCAAGAGACAGAAATCAGGA 1440

Qy   6741 ctaagtcctctgcttcagtttcattgttaacgggccttattctgatctcacctgtcgcgt 6800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCTTATTCTGATCTCACCTGTCGCGT 1500

Qy   6801 agctctaatattcacataaactgaaataaagaagtggaatgaggagctttg 6851
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCTCTAATATTCACATAAACTGAAATAAAGAAGTGGAATGAGGAGCTTTG 1551
```

```
RESULT   8
AF178030/c
LOCUS       AF178030   310002 bp    DNA             PRI       20-JAN-2000
DEFINITION  Homo sapiens chromosome 8 clone HPAC24 map 8q24.1, complete
            sequence.
ACCESSION   AF178030
VERSION     AF178030.1  GI:6539738
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 310002)
  AUTHORS   Momeni,P., Glockner,G., Schmidt,O., von Holtum,D., Albrecht,B.,
            Gillessen-Kaesbach,G., Hennekam,R., Meinecke,P., Zabel,B.,
            Rosenthal,A., Horsthemke,B. and Ludecke,H.J.
  TITLE     Mutations in a new gene, encoding a zinc-finger protein, cause
            tricho-rhino-phalangeal syndrome type I
  JOURNAL   Nat. Genet. 24 (1), 71-74 (2000)
  MEDLINE   20082812
   PUBMED   10615131
REFERENCE   2  (bases 1 to 310002)
  AUTHORS   Gloeckner,G., Rosenthal,A., Luedecke,H.-J. and Horsthemke,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .310002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24.1"
                     /clone="HPAC24"
     misc_feature    1. .117494
                     /note="Region: PAC22"
     repeat_region   complement(893. .1125)
                     /rpt_family="MIR"
     repeat_region   1549. .1608
                     /note="(GA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   2405. .2430
                     /rpt_type=tandem
     misc_feature    complement(2855. .2947)
                     /note="Data base match to GB|AA214657|AA214657"
     repeat_region   complement(2949. .3186)
                     /rpt_family="LTR43"
     repeat_region   4033. .4319
                     /rpt_family="L1MD3"
     repeat_region   4084. .4319
                     /rpt_family="L1MD2"
     repeat_region   4432. .4949
                     /rpt_family="MER2"
     repeat_region   4484. .4778
                     /rpt_family="AluY"
     repeat_region   4495. .4778
                     /rpt_type=inverted
     repeat_region   4990. .5085
                     /rpt_type=tandem
     repeat_region   5129. .5190
                     /rpt_family="MER2"
     repeat_region   5238. .5266
                     /note="(TA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   5240. .5269
                     /rpt_type=tandem
     repeat_region   complement(5323. .5609)
                     /rpt_family="AluY"
     repeat_region   5325. .5609
                     /rpt_type=inverted
     misc_feature    5361. .5488
                     /note="Region: GC Content > 60%"
     repeat_region   complement(5667. .5730)
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   complement(5828. .5940)
                     /rpt_family="MLT1I"
     repeat_region   complement(5950. .6039)
                     /rpt_family="MLT1I"
     repeat_region   complement(6208. .6296)
                     /note="(GA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   6215. .6296
                     /rpt_type=tandem
     repeat_region   6403. .6709
                     /rpt_family="MLT1D"
     repeat_region   complement(6716. .6756)
                     /rpt_family="MLT2CA"
     repeat_region   complement(6716. .6764)
                     /rpt_family="MLT2FA"
     repeat_region   6776. .6827
                     /note="(TAGA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   6801. .6856
                     /rpt_type=tandem
     repeat_region   complement(6828. .7342)
                     /rpt_family="MLT2D"
     repeat_region   7343. .7455
                     /rpt_family="MLT1D"
     repeat_region   complement(7470. .7791)
                     /rpt_family="MLT1J"
     repeat_region   complement(7796. .8404)
                     /rpt_family="MER41B"
     repeat_region   8416. .9267
                     /rpt_family="L1"
     repeat_region   complement(8418. .8867)
                     /rpt_family="L1PA5"
     repeat_region   complement(9268. .9319)
                     /rpt_family="MER41C"
     repeat_region   9416. .10052
                     /rpt_family="L2"
     repeat_region   9939. .10372
                     /rpt_family="L2a"
     repeat_region   complement(10386. .10430)
                     /rpt_family="MLT1J"
     repeat_region   complement(10497. .10549)
                     /rpt_family="MER5B"
     repeat_region   complement(11160. .11710)
                     /rpt_family="MER1A"
     repeat_region   11431. .11455
                     /rpt_type=inverted
     repeat_region   11510. .11533
                     /rpt_type=inverted
     misc_feature    complement(11731. .11894)
                     /note="Data base match to GB|AA803008|AA803008"
     misc_feature    12828. .13002
                     /note="Data base match to EMB|AU060784|AU060784"
     repeat_region   13185. .13292
                     /rpt_family="MER34"
     repeat_region   13204. .13351
                     /rpt_family="LTR29"
     repeat_region   13510. .13563
                     /rpt_type=tandem
     repeat_region   complement(13544. .13840)
                     /rpt_family="AluSg"
     repeat_region   13681. .13744
                     /rpt_type=tandem
     misc_feature    13879. .14015
                     /note="Data base match to GB|AA481774|AA481774"
     misc_feature    13893. .14014
                     /note="Data base match to GB|W26158|W26158"
     misc_feature    complement(13893. .14014)
                     /note="Data base match to GB|W28876|W28876"
     repeat_region   14016. .14123
                     /rpt_family="LTR29"
     misc_feature    14254. .14382
                     /note="Data base match to GB|AI571864|AI571864"
     repeat_region   complement(14407. .15223)
                     /rpt_family="MER31-internal"
     repeat_region   15301. .15344
                     /rpt_type=tandem
     repeat_region   complement(15578. .15853)
                     /rpt_family="L1M4_orf2"
     repeat_region   15859. .16212
                     /rpt_family="THE1B"
     repeat_region   complement(16229. .16303)
                     /rpt_family="L1M4_orf2"
```

```
     repeat_region   16304. .16630
                     /rpt_family="L2"
     repeat_region   complement(16797. .17093)
                     /rpt_family="AluSx"
     repeat_region   17396. .17500
                     /rpt_family="L2a"
     repeat_region   17813. .18145
                     /rpt_family="MLT1I"
     repeat_region   complement(18148. .18312)
                     /rpt_family="MIR"
     repeat_region   complement(19033. .19117)
                     /rpt_family="MER3"
     repeat_region   complement(19182. .19441)
                     /rpt_family="L2b"
     repeat_region   complement(19296. .19769)
                     /rpt_family="L2"
     repeat_region   complement(20088. .20276)
                     /rpt_family="MLT1A1"
     repeat_region   20260. .20315
                     /rpt_family="MLT1A1"
     repeat_region   20316. .20749
                     /rpt_family="L1PA5"
     repeat_region   21831. .22268
                     /rpt_family="LTR37B"
     repeat_region   22584. .22635
                     /rpt_type=tandem
     repeat_region   22961. .22999
                     /rpt_type=tandem
     repeat_region   complement(22970. .22992)
                     /note="(CAA)n"
                     /rpt_family="Simple_repeat"
     misc_feature    complement(24131. .24266)
                     /note="Data base match to GB|AA916901|AA916901"
     repeat_region   24971. .25498
                     /rpt_family="L2"
     repeat_region   complement(25501. .25640)
                     /rpt_family="MER45"
     repeat_region   25652. .27034
                     /rpt_family="L2"
     repeat_region   25703. .26136
                     /rpt_family="MSTB"
     repeat_region   26880. .27319
                     /rpt_family="L2a"
     misc_feature    28609. .28703
                     /note="Data base match to EMB|AU060784|AU060784"
     misc_feature    29402. .29582

  Query Match           11.3%;  Score 1127.4;  DB 88;  Length 310002;
  Best Local Similarity  99.9%;  Pred. No. 1e-189;
  Matches 6585;  Conservative    0;  Mismatches    1;  Indels    4;  Gaps    4;

Qy   3421 gactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattattaggaggag 3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36618 GACTCCCAGGCCTTTAAACATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAG 36559

Qy   3481 aacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagag 3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36558 AACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAG 36499

Qy   3541 gggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaagatcatct 3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36498 GGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCT 36439

Qy   3601 aactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaagcccaggg 3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36438 AACTGAAAGTCACCAGAGAGAAATTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGG 36379

Qy   3661 ttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactctggatat 3720
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  36378 TTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAGTCCTGGTCAGCCAAACTTTGGATAT 36319

Qy   3721 tcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagtactggaga 3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  36318 TCACAAAAGGATGCAACCTTTGCACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGA 36259

Qy   3781 tccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggcagtcctat 3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36258 TCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTAT 36199

Qy   3841 agaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagccctattga 3900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36198 AGAAAAGTACATGAGACCTGCGAAACACCCAAATTATTCACCACCAGGCAGCCCTATTGA 36139

Qy   3901 aaagtaccagtacccacttttttggacttcccttt-tacataatgacttccagagtgaagc 3959
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  36138 AAAGTACCAGTACCCACTTTTTGGACTTCCCTTTGTACATAATGACTTCCAGAGTGAAGC 36079

Qy   3960 tgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgcactactt 4019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36078 TGATTGGCTGCGGTTCTGGAGTAAATATAAGCTCTCCGTTCCTGGGAATCCGCACTACTT 36019

Qy   4020 gagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatcccaccttcaa 4079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36018 GAGTCACGTGCCTGGCCTACCAAATCCTTGCCAAAACTATGTGCCTTATCCCACCTTCAA 35959

Qy   4080 tctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatttggcgat 4139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35958 TCTGCCTCCTCATTTTTCAGCTGTTGGATCAGACAATGACATTCCTCTAGATTTGGCGAT 35899

Qy   4140 caagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacgaaggcacc 4199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35898 CAAGCATTCCAGACCTGGGCCAACTGCAAACGGTGCCTCCAAGGAGAAAACGAAGGCACC 35839

Qy   4200 accaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttgatagaag 4259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35838 ACCAAATGTAAAAAATGAAGGTCCCTTGAATGTAGTAAAAACAGAGAAAGTTGATAGAAG 35779

Qy   4260 tactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctggatgaagt 4319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35778 TACTCAAGATGAACTTTCAACAAAATGTGTGCACTGTGGCATTGTCTTTCTGGATGAAGT 35719

Qy   4320 gatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgcagcatatg 4379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35718 GATGTATGCTTTGCATATGAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATG 35659

Qy   4380 ccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcctgcatag 4439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35658 CCAGCATCTTTGCACGGACAAATATGACTTCACAACACATATCCAGAGGGGCCTGCATAG 35599

Qy   4440 gaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagcacttagca 4499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35598 GAACAATGCACAAGTGGAAAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCA 35539

Qy   4500 caattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaag 4559
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35538 CAATTAAATAGAAATAGGTTTTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAG 35479

Qy   4560 acctattaaaaaaatacttcatagagcctgccttatccaacatgaaattcccttcttttg 4619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35478 ACCTATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTTG 35419

Qy   4620 ttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatga 4679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35418 TTATTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGA 35359

Qy   4680 gaaagaatggaagatggtaaacaatcacttttttaaaacctgttaagtcaaaaccatcttg 4739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35358 GAAAGAATGGAAGATGGTAAACAATCACTTTTTTAAAACCTGTTAAGTCAAAACCATCTTG 35299

Qy   4740 gctaatatgtactggggaaataatccataagagatatcaccagactagaattaatatatt 4799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35298 GCTAATATGTACTGGGGAAATAATCCATAAGAGATATCACCAGACTAGAATTAATATATT 35239

Qy   4800 tataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatactaa 4859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35238 TATAAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGTTTACATATTATTATACTAA 35179
```

```
Qy   4860 agcagtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaat 4919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35178 AGCAGTACTGGACTGGCCATTGGACCATTTGTTCCAAAACCCATAAATTGTTGCCTAAAT 35119

Qy   4920 ttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaaa 4979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35118 TTATAATGATCATGAAACCCTAGGCAGAGGAGGAGAAATTGAAGGTCCAGGGCAATGAAA 35059

Qy   4980 gaaaaatggcgcccctctcaatttagtcttctctcattggccatgtttcagattttgacct 5039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35058 GAAAAATGGCGCCCTCTCAATTTAGTCTTCTCTCATTGGCCATGTTTCAGATTTTGACCT 34999

Qy   5040 agaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtgg 5099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34998 AGAAATGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTGG 34939

Qy   5100 cacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctgatattttt 5159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34938 CACGCTGTTTTTACCCAGCCCTGCCTGTACATACACATGCACACCCTCTCTGATATTTTT 34879

Qy   5160 gtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattcattttgtc 5219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34878 GTCCTTTAGATGTTCAAATACTCAGTAGTCCTTTTGTTTGCGGTTTAGATTCATTTTGTC 34819

Qy   5220 cacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgatttcatttta 5279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34818 CACACATGTACCCATTTTAAAAAACAATGTCCTCGATGCTTCTGTAGTGATTTCATTTTA 34759

Qy   5280 gccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaagcctcct 5339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34758 GCCAGGTATTTCTTTCTTGTGTGTGATGAACCAGTATGGATTTGCTTTTCTAAGCCTCCT 34699

Qy   5340 gttggttactaatctcacttggcacattataactaaaggaatcccctcaattcaaaagca 5399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34698 GTTGGTTACTAATCTCACTTGGCACATTATAACTAAAGGAATCCCCTCAATTCAAAAGCA 34639

Qy   5400 tagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttcttc 5459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34638 TAGATGGATACAAATGTCAGACCGTGGGTTTAATTTGTTTAGAACACATGGCATTTCTTC 34579

Qy   5460 acaaggtaacctgctgtatttatttatttttcttttggttaaatataaatttccaaacttttg 5519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34578 ACAAGGTAACCTGCTGTATTTATTTATTTTCTTTTGGTTAAATATAATTTCCAAACTTTG 34519

Qy   5520 tggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgtaccagcc 5579
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34518 TGGTCAGGCAGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAGCC 34459

Qy   5580 aatcccttcattaaatgtatacagatttagttaagtagcattaaataggattcttagaag 5639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34458 AATCCCTTCATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAG 34399

Qy   5640 tatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaagggaaag 5699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34398 TATGTCCTCATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAG 34339

Qy   5700 ctcctcagcataactgctcagggaaatagggctaaataactgaacattaaataattggtt 5759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34338 CTCCTCAGCATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGGTT 34279

Qy   5760 aaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactgact 5819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34278 AAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGACT 34219

Qy   5820 ttaataatttgcattatattgtcccaaccagtagtttatttttgccacggagatgtaga 5879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34218 TTAATAATTTGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTAGA 34159

Qy   5880 agatattacaagctactggatgcactgtcagattaacttatttcattaaagaagttggga 5939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34158 AGATATTACAAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGA 34099
```

```
Qy   5940 gaacaaatagg-aaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaa 5998
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34098 GAACAAATAGGAAAAAAAAAACTTATTTTTCTAGTAAATATTAATGTATTACATTTCAAA 34039

Qy   5999 taatggtgcctgacatattgaataattattttctacagtgtacgtatgcaacaaagatat 6058
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  34038 TAATGGTGCCTGACATATTGAATAATTATTTTCTACAGTGTACGTATGCAACAAAGATAT 33979

Qy   6059 tccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaaatgtagc 6118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33978 TCCATCATGCATTAGAGTCAGTTCTGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAGC 33919

Qy   6119 aaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtg 6178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33918 AAACAAGGTAATGAAGCAACTATTTCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGTG 33859

Qy   6179 tgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatgg 6238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33858 TGCATTAAAGTTGTAAACGGTAACATGAAACAAATGAAAGTTCTTGCTATAATGGTATGG 33799

Qy   6239 aaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcactta 6298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33798 AAAACAAGAAGGAAATGAAAATATTTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTTA 33739

Qy   6299 ttcattccaagtacttttttttttttttaatttttaagctcttaactcacattgttatgctt 6358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33738 TTCATTCCAAGTACTTTTTTTTTTTTTAATTTTTAAGCTCTTAACTCACATTGTTATGCTT 33679

Qy   6359 aagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaaacatttc 6418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33678 AAGATGATAAACATATATCCTCTTTTTATTGCTTTGTCTATGTTTCATATGAAACATTTC 33619

Qy   6419 agaaattatttttgataagtgttgctggaatctgcaacgctgatttttttttgcattctgt 6478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33618 AGAAATTATTTTGATAAGTGTTGCTGGAATCTGCAACGCTGATTTTTTTTTGCATTCTGT 33559

Qy   6479 agtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattgttttgtt 6538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33558 AGTCGCATTTGCACTCCATTTTTACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGTT 33499

Qy   6539 tgggttttgtttcttttttcacagtgccgggtcttcgtttcttaaagttggatggcaggta 6598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33498 TGGGTTTTGTTTCTTTTTCACAGTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGTA 33439

Qy   6599 gagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatgt 6658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33438 GAGTTCAACCAGTTCGTGACTGTTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATGT 33379

Qy   6659 tgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagaga 6718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33378 TGTGTTGTCATTTTCATTTTTGCATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAGA 33319

Qy   6719 aagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaacgggccttt 6778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33318 AAGCAAGAGACAGAAATCAGGACTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCTT 33259

Qy   6779 attctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaagaagtgga 6838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33258 ATTCTGATCTCACCTGTCGCGTAGCTCTAATATTCACATAAACTGAAATAAAGAAGTGGA 33199

Qy   6839 atgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaattttgatg 6898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33198 ATGAGGAGCTTTGACATTCAAATTATGTGATGTAATTTATCTTCCTTAGGAATTTTGATG 33139

Qy   6899 gatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaaagag 6958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33138 GATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAAAGAG 33079

Qy   6959 aggagattccccccaaacaacaatatttaattttcttagtaaaaagaataacagaatgcat 7018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33078 AGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAATGCAT 33019

Qy   7019 cgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacaccagaa 7078
```

```
Db 33018 CGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACCAGAA 32959

Qy  7079 gtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactcattaagc 7138
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32958 GTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATTAAGC 32899

Qy  7139 aatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttta 7198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32898 AATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTTTTTA 32839

Qy  7199 ttttcattcaatttttttttttttctcagacagacatagtagtatcaactagcattggaaa 7258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32838 TTTTCATTCAATTTTTTTTTTTTTCTCAGACAGACATAGTAGTATCAACTAGCATTGGAAA 32779

Qy  7259 atacatatcactattcttggaatatttatggtcagtctacttttttagtaaaatatttttg 7318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32778 ATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTTAGTAAAATATTTTTG 32719

Qy  7319 gatagcgttgacacgatagatcttattccatacttctttattattgataatttttattttc 7378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32718 GATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTTC 32659

Qy  7379 atttttttgctttcattattatacatattttggtggagaagaggttgggctttttttgaaag 7438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32658 ATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTTGAAAG 32599

Qy  7439 agacaaaaatttattataacactaaacactccttttttgacatattaaagccttttattcc 7498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32598 AGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTCC 32539

Qy  7499 atctctcaagatatattataaaatttatttttttaatttaagatttctgaattattttat 7558
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32538 ATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATTTTAT 32479

Qy  7559 cttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaatttc 7618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32478 CTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTTAATGAGGAATTTC 32419

Qy  7619 atgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgc 7678
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32418 ATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTGC 32359

Qy  7679 cagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatattc 7738
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32358 CAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATTC 32299

Qy  7739 taaattgagctttggatcaaacttttaggccaggaccagctcatgcgttctcattcttcct 7798
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32298 TAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCCT 32239

Qy  7799 tttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatagaaaa 7858
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32238 TTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAAA 32179

Qy  7859 atcataaagagccaacccatctcagaacgttgtggattgagagagacactacatgactcc 7918
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32178 ATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTCC 32119

Qy  7919 aagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg-aaagag 7977
         |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db 32118 AAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAAAGAG 32059

Qy  7978 tatgcccaattctctctacatgacatattgagatttttttaatcaacttttaagatagt 8037
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 32058 TATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGATAGT 31999

Qy  8038 gatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagcatgg 8097
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31998 GATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATGG 31939

Qy  8098 ggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctctttg 8157
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 31938 GGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTTG 31879

Qy  8158 ttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctattatga 8217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31878 TTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATGA 31819

Qy  8218 tccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctgagga 8277
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31818 TCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGGA 31759

Qy  8278 aatacatttttcttaataaactgaagagtctcagtatggctattaaaataattattagcc 8337
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31758 AATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGCC 31699

Qy  8338 tcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgctg 8397
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31698 TCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCTG 31639

Qy  8398 ccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccagtggt 8457
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31638 CCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGGT 31579

Qy  8458 ggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgcct 8517
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31578 GGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCT 31519

Qy  8518 taagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcaaaaaaa 8577
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31518 TAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAAAAAA 31459

Qy  8578 aaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgctgact 8637
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31458 AAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGCTGACT 31399

Qy  8638 gtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaaatat 8697
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31398 GTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTAAATAT 31339

Qy  8698 tctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatgagaa 8757
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31338 TCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAATGAGAA 31279

Qy  8758 gaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaagaaagac 8817
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31278 GAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCATTGTTATTTTCCCAAAGAAAGAC 31219

Qy  8818 ggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagtactgt 8877
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31218 GGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACTGT 31159

Qy  8878 ttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcatcccca 8937
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31158 TTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCCCA 31099

Qy  8938 gttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagtatgtc 8997
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31098 GTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGTC 31039

Qy  8998 acatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgtacat 9057
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 31038 ACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATGTACAT 30979

Qy  9058 tcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacaccggag 9117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 30978 TCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGGAG 30919

Qy  9118 aaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaaattgtcc 9177
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 30918 AAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTATCAGAGAAAATTGTCC 30859

Qy  9178 ttctagagcatagagtcccaaaactcaattctggttttccccctgttttttttttttttt 9237
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 30858 TTCTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCCTGTTTTTTTTTTTTTTT 30799
```

```
Qy   9238 ttttcccaacatatgaactgcagcatatcacttttttcttttttgtgcctcaggttcctcac 9297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30798 TTTTCCCAACATATGAACTGCAGCATATCACTTTTTCTTTTTTGTGCCTCAGGTTCCTCAC 30739

Qy   9298 ctgtaaaattgaaaaatatatgtattaataatattattaataataataatggtaatgtag 9357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30738 CTGTAAAATTGAAAAATATATGTATTAATAATATTATTAATAATAATAATGGTAATGTAG 30679

Qy   9358 tacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtgccaagta 9417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30678 TACTTGTTTGTAAAGCACTTTGAGATCCTTGGTTGAAAGGCACCATAGGAGTGCCAAGTA 30619

Qy   9418 ttattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaaaggttgg 9477
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30618 TTATTATGTGGCCAAGGGGTTATTTAAACTGTCAGTTCCCAAAGGCCAGGAAAGGTTGG 30559

Qy   9478 ggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgttgactat 9537
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30558 GGTCATTTTTCTTAAAGACGAGCTGTAAATATCAACTAGGCAGCCAATAGTGTTGACTAT 30499

Qy   9538 gaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctaccaataag 9597
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30498 GAAGATGCAAAACTATTACTAGGCTGATAAAATCATAGTTTCTTAATGGCTACCAATAAG 30439

Qy   9598 gcaaatatcacaataataaacgccaaattccttagggcggactatttgacaaccacatgg 9657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30438 GCAAATATCACAATAATAAACGCCAAATTCCTTAGGGCGGACTATTTGACAACCACATGG 30379

Qy   9658 aaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaatttaactt 9717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30378 AAAACTTTGGGGGAGGCATGAGGGGGGAACATCTCAAAATGCCAATGTAAAATTTAACTT 30319

Qy   9718 acagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttgctaagaa 9777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30318 ACAGCAATATTCACCAGCAGAAAATGTCTTTCATATGGAATGATTTCATGTTGCTAAGAA 30259

Qy   9778 aaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatagttctgtt 9837
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30258 AAAGAATTCAATTTGTAGTCCTGATTTGAATACTAGAATGTTGGCTATAATAGTTCTGTT 30199

Qy   9838 cttacaacacatgaaatttttcgttttatttttatttgttttcatagtgcatgttcatt 9897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30198 CTTACAACACATGAAATTTTTTCGTTTTATTTTATTTTGTTTTCATAGTGCATGTTCATT 30139

Qy   9898 tctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggcagcaaaga 9957
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30138 TCTACTCACAAACATGTTCTTGGTGTATTTCTTATGCAAACAATCTTCAGGCAGCAAAGA 30079

Qy   9958 tgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||| |||||||||||||||
Db  30078 TGTCTGTTACATCTAAACTTGAATAATAAAGTTTTACCACCAGTTACACA 30029


RESULT   9
AX067304
LOCUS       AX067304      996 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 8 from Patent WO0078960.
ACCESSION   AX067304
VERSION     AX067304.1  GI:12544928
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   Yuqiu,J. and Mitcham,J.L.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0078960-A 8 28-DEC-2000;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .996
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      299 a    148 c    188 g    361 t
ORIGIN


  Query Match             9.3%;  Score 931.4;  DB 10;  Length 996;
  Best Local Similarity   99.6%;  Pred. No. 6e-155;
  Matches  965;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy   5891 gctactggatgcactgtcagattaacttatttcattaaagaagttgggagaacaaatagg 5950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     29 GCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGAGAACAAATAGG 88

Qy   5951 -aaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaataatggtgcct 6009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 AAAAAAAAAACTTATTTTTCTAGTAAATATTAATGTATTACATTTCAAATAATGGTGCCT 148

Qy   6010 gacatattgaataattattttctacagtgtacgtatgcaacaaagatattccatcatgca 6069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 GACATATTGAATAATTATTTTCTACAGTGTACGTATGCAACAAAGATATTCCATCATGCA 208

Qy   6070 ttagagtcagttctggctctgcctagctgtttacatttgcaaatgtagcaaacaaggtaa 6129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 TTAGAGTCAGTTCTGGCTCTGCCTAGCTGTTTACATTTGCAAATGTAGCAAACAAGGTAA 268

Qy   6130 tgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtgtgcattaaagt 6189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 TGAAGCAACTATTTCTATTGCAGTAGATATCCTTTTGTGTGTGTGTGTGTGCATTAAAGT 328

Qy   6190 tgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatggaaaacaagaag 6249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 TGTAAACGGTAACATGAAACAAATGAAAGTTCTTGCTATAATGGTATGGAAAACAAGAAG 388

Qy   6250 gaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcacttattcattccaag 6309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 GAAATGAAAATATTTTTATGCCTACTTAGGAAAAAAAGGGTAGCACTTATTCATTCCAAG 448

Qy   6310 tacttttttttttttaatttttaagctcttaactcacattgttatgcttaagatgataaa 6369
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 TAC-TTTTTTTTTTAATTTTTAAGCTCTTAACTCACATTGTTATGCTTAAGATGATAAA 507

Qy   6370 catatatcctcttttttattgctttgtctatgtttcatatgaaacatttcagaaattattt 6429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    508 CATATATCCTCTTTTTATTGCTTTGTCTATGTTTCATATGAAACATTTCAGAAATTATTT 567

Qy   6430 tgataagtgttgctggaatctgcaacgctgat-tttttttgcattctgtagtcgcattt 6488
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    568 TGATAAGTGTTGCTGGAATCTGCAACGCTGATGTTTTTTTTGCATTCTGTAGTCGCATTT 627

Qy   6489 gcactccatttttacattaattcgcagttgctttgtatcattgttttgtttgggttttgt 6548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    628 GCACTCCATTTTTACATTAATTCGCAGTTGCTTTGTATCATTGTTTTGTTTGGGTTTTGT 687

Qy   6549 ttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggtagagttcaacc 6608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688 TTCTTTTTCACAGTGCCGGGTCTTCGTTTCTTAAAGTTGGATGGCAGGTAGAGTTCAACC 747

Qy   6609 agttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatgttgtgttgtca 6668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    748 AGTTCGTGACTGTTGTAGCGAATGAAGTTAAAAAAATGTCTTTCTGATGTTGTGTTGTCA 807

Qy   6669 ttttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagagaaagcaagaga 6728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    808 TTTTCATTTTTGCATTTTTTTGTTTGCATATTAAAAAAAGAGAAAAGAGAAAGCAAGAGA 867

Qy   6729 cagaaatcaggactaagtcctctgcttcagtttcattgttaacgggccttattctgatct 6788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    868 CAGAAATCAGGACTAAGTCCTCTGCTTCAGTTTCATTGTTAACGGGCCTTATTCTGATCT 927

Qy   6789 cacctgtcgcgtagctctaatattcacataaactgaaataaagaagtggaatgaggagct 6848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    928 CACCTGTCGCGTAGCTCTAATATTCACATAAACTGAAATAAAGAAGTGGAATGAGGAGCT 987
```

```
Qy   6849 ttgacattc 6857
          |||||| ||
Db    988 TTGACAGTC 996


RESULT 10
AC013815/c
LOCUS       AC013815   167123 bp    DNA             HTG       07-JUN-2000
DEFINITION  Homo sapiens clone RP11-21J16, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC013815
VERSION     AC013815.4  GI:8316201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 7, 2000 this sequence version replaced gi:7341965.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4033
                Center clone name: 21_J_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160651 bases at least Q40
                Consensus quality: 163054 bases at least Q30
                Consensus quality: 164289 bases at least Q20
                Insert size: 167000; agarose-fp
                Insert size: 166123; sum-of-contigs
                Quality coverage: 6.6 in Q20 bases; agarose-fp
                Quality coverage: 6.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1116: contig of 1116 bp in length
            *     1117 1216: gap of      100 bp
            *     1217     2561: contig of 1345 bp in length
            *     2562 2661: gap of      100 bp
            *     2662     3754: contig of 1093 bp in length
            *     3755 3854: gap of      100 bp
            *     3855     4514: contig of 660 bp in length
            *     4515 4614: gap of      100 bp
            *     4615    11771: contig of 7157 bp in length
            *    11772 11871: gap of      100 bp
            *    11872    22787: contig of 10916 bp in length
            *    22788 22887: gap of      100 bp
            *    22888    31860: contig of 8973 bp in length
            *    31861 31960: gap of      100 bp
            *    31961    53736: contig of 21776 bp in length
            *    53737 53836: gap of      100 bp
            *    53837    81814: contig of 27978 bp in length
            *    81815 81914: gap of      100 bp
            *    81915   123726: contig of 41812 bp in length
            *   123727 123826: gap of      100 bp
            *   123827   167123: contig of 43297 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167123
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21J16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1116
                     /note="assembly_fragment"
     misc_feature    1217. .2561
                     /note="assembly_fragment"
     misc_feature    2662. .3754
                     /note="assembly_fragment"
     misc_feature    3855. .4514
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    4615. .11771
                     /note="assembly_fragment"
     misc_feature    11872. .22787
                     /note="assembly_fragment"
     misc_feature    22888. .31860
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    31961. .53736
                     /note="assembly_fragment"
     misc_feature    53837. .81814
                     /note="assembly_fragment"
     misc_feature    81915. .123726
                     /note="assembly_fragment"
     misc_feature    123827. .167123
                     /note="assembly_fragment"
BASE COUNT    51442 a  29991 c  31207 g  53480 t   1003 others
ORIGIN


  Query Match             8.3%;  Score 827.2;  DB 63;  Length 167123;
  Best Local Similarity   93.4%;  Pred. No. 1.3e-136;
  Matches  889;  Conservative    0;  Mismatches   53;  Indels   10;  Gaps    2;

Qy    9065 ttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacaccggagaaaagca 9124
           || | || |||| ||| ||| ||   |||| ||   |   |||| | ||||| ||||  |
Db   22782 TTTTTTTTAAAGAAAAAGGAAAGTGTTAAGTTTTCTTTGTTCCTCCCCCGGAAAAAAAAA 22723

Qy    9125 ggaataactttaccgtggaaataatgtttagcttttatcagagaaaattgtccttctaga 9184
           |||| | |||| || |||||| || |||||||||| | | | |||||   |||| || |
Db   22722 GGAAAACCTTTCCCCTGGAAAAAAGGTTTAGCTTTTTTAAAAAAAAATGTCCCTTTTAAA 22663

Qy    9185 gcatagagtcccaaaactcaattctggttttcccc---------tgtttttttttttttt 9235
           | | |  |||||||||  |||||  || ||| |||         | ||||||||||||||
Db   22662 GAAAAAGGTCCCAAAAACCAATTTGGGGTTTTCCCCCGGTTTTTTTTTTTTTTTTTTTT 22603

Qy    9236 ttttttcccaacatatgaactgcagcatatcacttttttctttttgtgcctcaggttcctc 9295
```

```
Db  22602 TTTTTTCCCAACATATGAACTGCAGCATATCACTTTTTCTTTTTGTGCCTCAGGTTCCTC 22543

Qy   9296 acctgtaaaattgaaaaatatatgtattaataatattattaataataatgtggtaatgt 9355
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22542 AGCTGTAAAATTGAAAAATATATGTATTAATAATATTATTAATAATAATAATGGTAATGT 22483

Qy   9356 agtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtgccaag 9415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22482 AGTACTTGTTTGTAAAGCACTTTGAGATCCTTGGTTGAAAGGCACCATAGGAGTGCCAAG 22423

Qy   9416 tattattatgtggccaaggggggttatttaaactgtcagttcccaaaggccaggaaaggtt 9475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22422 TATTATTATGTGGCCAAGGGGGTTATTTAAACTGTCAGTTCCCAAAGGCCAGGAAAGGTT 22363

Qy   9476 ggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgttgact 9535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22362 GGGGTCATTTTTCTTAAAGACGAGCTGTAAATATCAACTAGGCAGCCAATAGTGTTGACT 22303

Qy   9536 atgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctaccaata 9595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22302 ATGAAGATGCAAAACTATTACTAGGCTGATAAAATCATAGTTTCTTAATGGCTACCAATA 22243

Qy   9596 aggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaaccacat 9655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22242 AGGCAAATATCACAATAATAAACGCCAAATTCCTTAGGGCGGACTATTTGACAACCACAT 22183

Qy   9656 ggaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaatttaac 9715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22182 GGAAAACTTTGGGGGAGGCATGAGGGGGGAACATCTCAAAATGCCAATGTAAAATTTAAC 22123

Qy   9716 ttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttgctaag 9775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22122 TTACAGCAATATTCACCAGCAGAAAATGTCTTTCATATGGAATGATTTCATGTTGCTAAG 22063

Qy   9776 aaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatagttctg 9835
          |||||||||||||||||| |||||||||||||||| ||||||||||||||||||||||
Db  22062 AAAAAGAATTCAATTTGTAGTCTTGATTTGAATACTAGAATGTGGGCTATAATAGTTCTG 22003

Qy   9836 ttcttacaacacatgaaatttttttcgttttatttttatttttgttttcatagtgcatgttca 9895
          |||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  22002 TTCTCACAACACATGAAATTTTTTCGTTTTATTTTATTTTGTTTTCATAGAGCATGTTCA 21943

Qy   9896 tttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggcagcaaa 9955
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  21942 TTTCTACTCACAAACATGTTCTTGGTGTATTTTTTATGCAAACAATCTTCAGGCAGCAAA 21883

Qy   9956 gatgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||||||| |||||||||||||
Db  21882 GATGTCTGTTACATCTAAACTTGAATAATAAAGTTTTACCACCAGTTACACA 21831


RESULT  11
AX067327
LOCUS       AX067327      674 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 31 from Patent WO0078960.
ACCESSION   AX067327
VERSION     AX067327.1  GI:12544951
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Yuqiu,J. and Mitcham,J.L.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0078960-A 31 28-DEC-2000;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

BASE COUNT      222 a    110 c    128 g    213 t      1 others
ORIGIN


  Query Match             6.7%;  Score 668.8;  DB 10;  Length 674;
  Best Local Similarity   99.4%;  Pred. No. 1.7e-108;
  Matches  670;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   8958 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 9017
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGTCACATGAAGAAAATGTGAAAT 60

Qy   9018 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 9077
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCAAGAGTAATCCACACGTGAGAAACTAGACAATGTACATTCATGTGTTCTCTTGAAAGG 120

Qy   9078 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 9137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGGAGAAAAGCAGGAATAACTTTAC 180

Qy   9138 cgtggaaataatgtttagcttttatcagagaaaattgtccttctagagcatagagtccca 9197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGTGGAAATAATGTTTAGCTTTTATCAGAGAAAATTGTCCTTCTAGAGCATAGAGTCCCA 240

Qy   9198 aaactcaattctggttttccccctgttttttttttttttttttttcccaacatatgaactg 9257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAACTCAATTCTGGTTTTCCCCTGTTTTTTTTTTTTTTTTTTTCCCAACATATGAACTG 300

Qy   9258 cagcatatcacttttttctttttgtgcctcaggttcctcacctgtaaaattgaaaaatata 9317
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    301 CAGCATATCACTTTTTCTTTTTGTGCCTCAGGTTCCTCAGCTGTAAAATTGAAAAATATA 360

Qy   9318 tgtattaataatattattaataataatgtaatgtagtacttgtttgtaaagcactt 9377
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTATTAATAATATTATTAATAATAATAATGGTAATGTAGTACTTGTTTGTAAAGCACTT 420

Qy   9378 tgagatccttggttgaaaggcaccataggagtgccaagtattattatgtggccaaggggg 9437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGAGATCCTTGGTTGAAAGGCACCATAGGAGTGCCAAGTATTATTATGTGGCCAAGGGGG 480

Qy   9438 ttatttaaactgtcagttcccaaaggccaggaaaggttggggtcatttttcttaaagacg 9497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTATTTAAACTGTCAGTTCCCAAAGGCCAGGAAAGGTTGGGGTCATTTTTCTTAAAGACG 540

Qy   9498 agctgtaaatatcaactaggcagccaatagtgttgactatgaagatgcaaaactattact 9557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCTGTAAATATCAACTAGGCAGCCAATAGTGTTGACTATGAAGATGCAAAACTATTACT 600

Qy   9558 aggctgataaaatcatagtttcttaatggctaccaataaggcaaatatcacaataataaa 9617
          |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGGMTGATAAAATCATAGTTTCTTAATGGCTACCAATAAGGCAAATATCACAATAATAAA 660

Qy   9618 cgccaaattcctta 9631
          | ||||||||| ||
Db    661 CACCAAATTCCGTA 674


RESULT  12
AF130342/c
LOCUS       AF130342   350000 bp    DNA             PRI       09-JUL-1999
DEFINITION  Homo sapiens chromosome 8 clone PAC 87.1 map 8q24.1, complete
            sequence.
ACCESSION   AF130342
VERSION     AF130342.1  GI:5430714
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 350000)
  AUTHORS   Gloeckner,G., Rosenthal,A., Luedecke,H.-J. and Horsthemke,B.
  TITLE     Annotation of 642 kb of human genomic sequence in 8q24.1 Segment 1
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 350000)
  AUTHORS   Gloeckner,G., Rosenthal,A., Luedecke,H.-J. and Horsthemke,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .350000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24.1"
                     /clone="PAC 87.1"
     misc_feature    1. .145388
                     /note="Region: M1557Q13"
     misc_feature    2214. .2237
                     /note="195, (TAAAA)n, Simple_repeat, pos., in, rep.:,
                     4. .27"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     repeat_region   3095. .3214
                     /note="62.50, %, 24, length:, 6, #:, 20, seq:, atataa"
                     /rpt_type=tandem
                     /rpt_unit-score:
                     /evidence=not_experimental
     exon            3469. .3917
                     /note="Alignment: , 50.0, cds|PFARPI_1|PFARPI_1:, EST+CDS"
                     /evidence=not_experimental
     exon            3469. .3917
                     /note="Alignment: CDS, 50.0, cds|PFARPI_1|PFARPI_1:,
                     non-human"
                     /evidence=not_experimental
     misc_feature    3754. .3798
                     /note="39, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .45"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     exon            4576. .4774
                     /note="Alignment: CDS, 50.0, cds|PFARPI_1|PFARPI_1:,
                     non-human"
                     /evidence=not_experimental
     exon            4576. .4774
                     /note="Alignment: , 50.0, cds|PFARPI_1|PFARPI_1:, EST+CDS"
                     /evidence=not_experimental
     misc_feature    5534. .5562
                     /note="23, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .29"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     exon            5757. .6161
                     /note="Alignment: EST, 50.0, gb|AI115824|AI115824:,
                     non-human"
                     /evidence=not_experimental
     exon            5757. .6161
                     /note="Alignment: , 50.0, gb|AI115824|AI115824:, EST+CDS"
                     /evidence=not_experimental
     misc_feature    6295. .6321
                     /note="21, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .27"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    7058. .7080
                     /note="23, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .23"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    7749. .7790
                     /note="212, (GAAA)n, Simple_repeat, pos., in, rep.:,
                     2. .41"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    7876. .7911
                     /note="24, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .36"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    8089. .8116
                     /note="220, POLY_A, Simple_repeat, pos., in, rep.:, 1. .28"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    10017. .10045
                     /note="23, AT_rich, Low_complexity, pos., in, rep.:,
                     1. .29"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    11839. .12361
                     /note="0.65, GC=, 63.30%, #37"
                     /note="Region: CpG Island"
                     /evidence=not_experimental
     misc_feature    11883. .12436
                     /note="21.60, 554, bp"
                     /note="Region: GC Content"
                     /evidence=not_experimental
     exon            11933. .12179
                     /note="Alignment: , 50.0, gb|AA958752|AA958752:, EST+CDS"
                     /evidence=not_experimental
     exon            11933. .12179
                     /note="Alignment: EST, 50.0, gb|AA958752|AA958752:,
                     non-human"
                     /evidence=not_experimental
     misc_feature    complement(11933. .11997)
                     /note="255, 1:HPAC87.X.1.00_Ex_4, 202. .266, of,
                     gb|AA958752|AA958752, ua17f06.r1, Soares, mouse, mammary,
                     gland, NbMMG, Mus, musculus, P = 3.2e-05, S = 255"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    12304. .12324
                     /note="21, GC_rich, Low_complexity, pos., in, rep.:,
                     1. .21"
                     /note="Region: Simple Repeats"
                     /evidence=not_experimental
     misc_feature    12559. .12654
                     /note="232, 1:HPAC87.X.1.00_Ex_6, 64546. .64641, of,
                     gb|AC004910|AC004910, Homo, sapiens, PAC, clone,
                     DJ0859M06, from, 7q11,, complete, P = 0.00058, S = 232"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    12559. .12654
                     /note="232, 2:HPAC87.X.1.00_Ex_6, 64541. .64636, of,
                     gb|AC004910|AC004910, Homo, sapiens, clone, DJ0859M06;,
                     HTGS, phase, 1,, 1, unordered, P = 0.00032, S = 232"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    complement(12569. .12653)
                     /note="205, 4:HPAC87.X.1.00_Ex_6, 194. .278, of,
                     gb|AI051595|AI051595, oz02b04.x1,
                     Soares_fetal_liver_spleen_1NFLS_S1, Homo, P = 0.00099, S =
                     205"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    complement(12569. .12653)
                     /note="205, 2:HPAC87.X.1.00_Ex_6, 194. .278, of,
                     gb|AI051595|AI051595, oz02b04.x1,
                     Soares_fetal_liver_spleen_1NFLS_S1, Homo, P = 0.0044, S =
                     205"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    complement(12569. .12653)
                     /note="186, 4:HPAC87.X.1.00_Ex_6, 67138. .67222, of,
                     gb|AF064860|AF064860, Homo, sapiens, chromosome, 21q22.3,
                     PAC, 70I24,, complete, P = 0.067, S = 186"
                     /note="Region: Data base match"
                     /evidence=not_experimental
     misc_feature    complement(12570. .12653)
                     /note="220, 1:HPAC87.X.1.00_Ex_6, 151. .234, of,
                     gb|AI053773|AI053773, qi73h01.x1, NCI_CGAP_Ov26, Homo,
                     sapiens, cDNA, clone, P = 0.0011, S = 220"
```

```
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   complement(12570. .12653)
                    /note="220, 1:HPAC87.X.1.00_Ex_6, 151. .234, of,
                    gb|AI053773|AI053773, qi73h01.x1, NCI_CGAP_Ov26, Homo,
                    sapiens, cDNA, clone, P = 0.00025, S = 220"
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   12571. .12653
                    /note="303, polypurine, Low_complexity, pos., in, rep.:,
                    6. .88"
                    /note="Region: Simple Repeats"
                    /evidence=not_experimental
     misc_feature   12571. .12653
                    /note="207, 3:HPAC87.X.1.00_Ex_6, 98. .180, of,
                    gb|AA537041|AA537041, vj87a09.r1, Knowles, Solter, mouse,
                    2, cell, Mus, musculus, cDNA, P = 0.0038, S = 207"
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   12571. .12641
                    /note="147, 3:HPAC87.X.1.00_Ex_6, 54144. .54214, of,
                    gb|AC004910|AC004910, Homo, sapiens, clone, DJ0859M06;,
                    HTGS, phase, 1,, 1, unordered, P = 0.89, S = 147"
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   12571. .12641.
                    /note="147, 2:HPAC87.X.1.00_Ex_6, 54149. .54219, of,
                    gb|AC004910|AC004910, Homo, sapiens, PAC, clone,
                    DJ0859M06, from, 7q11,, complete, P = 0.98, S = 147"
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   12572. .12654
                    /note="200, 5:HPAC87.X.1.00_Ex_6, 33. .115, of,
                    gb|AA896722|AA896722, vy09c04.r1, Stratagene, mouse,
                    macrophage, (#937306), Mus, P = 0.0071, S = 200"
                    /note="Region: Data base match"
                    /evidence=not_experimental
     misc_feature   12572. .12653
                    /note="190, 5:HPAC87.X.1.00_Ex_6, 70219. .70300, of,
                    gb|AC005189|AC005189, Homo, sapiens, PAC, clone,

  Query Match             4.6%;  Score 460.8;  DB 88;  Length 350000;
  Best Local Similarity   98.4%;  Pred. No. 8.4e-72;
  Matches  478;  Conservative    0;  Mismatches    2;  Indels    6;  Gaps    1;

Qy       1 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13394 TTCCTCCGCGAAGGCTCCTTTGATATTAATAGTGTTGGTGTCTTGAAACTGACGTAATGC 13335

Qy      61 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13334 GCGGAGACTGAGGTCCTGACAAGCGATAACATTTCTGATAAAGACCCGATCTTACTGCAA 13275

Qy     121 tctctagcgtcctctttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13274 TCTCTAGCGTCCTCTTTTTTGGTGCTGCTGGTTTCTCCAGACCTCGCGTCCTCTCGATTG 13215

Qy     181 ctctctcgccttcctatttctttttttttttttaaacaaaaaacaacacccccctcccct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13214 CTCTCTCGCCTTCCTATTTCTTTTTTTTTTTTTTAAACAAAAAACAACACCCCCTCCCCT 13155

Qy     241 ctcccacccggcaccgggcacatccttgctctatttcctttctctt------tctctctc 294
           ||||||||||||||||||||||||||||||||||||||||||||||      ||||||||
Db   13154 CTCCCACCCGGCACCGGGCACATCCTTGCTCTATTTCCTTTCTCTTTCTCTCTCTCTCTC 13095

Qy     295 tctctctctctttttttaataagggtgggggagggaaaggggggggaggcaggaaagacct 354
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   13094 TCTCTCTCTCTTTTTTAATAAGGGTGGGGGAGGGAAAGGGGGGGGATGCAGGAAAGACCT 13035

Qy     355 ttttctctccccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcat 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13034 TTTTCTCTCCCCCCCCGCAATAATCCAAGATCAACTCTGCAAACAACAGAAGACGGTTCAT 12975

Qy     415 ggctttggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatca 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12974 GGCTTTGGCCGCCGCGCCACCATCTTTCGGGCTGCCGAGGGTGTTCTTGACGATTAATCA 12915

Qy     475 acagat 480
           |||| |
Db   12914 ACAGGT 12909


RESULT  13
AC084171/c
LOCUS       AC084171    50698 bp    DNA             HTG       14-OCT-2000
DEFINITION  Homo sapiens chromosome 8 clone CTD-2183P16 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC084171
VERSION     AC084171.1  GI:10801399
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50698)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone CTD-2183P16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 50698)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11361
                Center clone name: 2183_P_16
            --------------.
            * NOTE: This record contains 60 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      722: contig of 722 bp in length
```

```
*      723 822: gap of      100 bp
*      823     1604: contig of 782 bp in length
*     1605 1704: gap of      100 bp
*     1705     2441: contig of 737 bp in length
*     2442 2541: gap of      100 bp
*     2542     3259: contig of 718 bp in length
*     3260 3359: gap of      100 bp
*     3360     4075: contig of 716 bp in length
*     4076 4175: gap of      100 bp
*     4176     4942: contig of 767 bp in length
*     4943 5042: gap of      100 bp
*     5043     5809: contig of 767 bp in length
*     5810 5909: gap of      100 bp
*     5910     6689: contig of 780 bp in length
*     6690 6789: gap of      100 bp
*     6790     7553: contig of 764 bp in length
*     7554 7653: gap of      100 bp
*     7654     8393: contig of 740 bp in length
*     8394 8493: gap of      100 bp
*     8494     9246: contig of 753 bp in length
*     9247 9346: gap of      100 bp
*     9347    10104: contig of 758 bp in length
*    10105 10204: gap of      100 bp
*    10205    10981: contig of 777 bp in length
*    10982 11081: gap of      100 bp
*    11082    11865: contig of 784 bp in length
*    11866 11965: gap of      100 bp
*    11966    12678: contig of 713 bp in length
*    12679 12778: gap of      100 bp
*    12779    13528: contig of 750 bp in length
*    13529 13628: gap of      100 bp
*    13629    14359: contig of 731 bp in length
*    14360 14459: gap of      100 bp
*    14460    15214: contig of 755 bp in length
*    15215 15314: gap of      100 bp
*    15315    16014: contig of 700 bp in length
*    16015 16114: gap of      100 bp
*    16115    16847: contig of 733 bp in length
*    16848 16947: gap of      100 bp
*    16948    17700: contig of 753 bp in length
*    17701 17800: gap of      100 bp
*    17801    18577: contig of 777 bp in length
*    18578 18677: gap of      100 bp
*    18678    19409: contig of 732 bp in length
*    19410 19509: gap of      100 bp
*    19510    20244: contig of 735 bp in length
*    20245 20344: gap of      100 bp
*    20345    21094: contig of 750 bp in length
*    21095 21194: gap of      100 bp
*    21195    21914: contig of 720 bp in length
*    21915 22014: gap of      100 bp
*    22015    22753: contig of 739 bp in length
*    22754 22853: gap of      100 bp
*    22854    23610: contig of 757 bp in length
*    23611 23710: gap of      100 bp
*    23711    24455: contig of 745 bp in length
*    24456 24555: gap of      100 bp
*    24556    25306: contig of 751 bp in length
*    25307 25406: gap of      100 bp
*    25407    26123: contig of 717 bp in length
*    26124 26223: gap of      100 bp
*    26224    27001: contig of 778 bp in length
*    27002 27101: gap of      100 bp
*    27102    27822: contig of 721 bp in length
*    27823 27922: gap of      100 bp
*    27923    28665: contig of 743 bp in length
*    28666 28765: gap of      100 bp
*    28766    29505: contig of 740 bp in length
*    29506 29605: gap of      100 bp
*    29606    30339: contig of 734 bp in length
*    30340 30439: gap of      100 bp
*    30440    31185: contig of 746 bp in length
*    31186 31285: gap of      100 bp
*    31286    32024: contig of 739 bp in length
*    32025 32124: gap of      100 bp
*    32125    32910: contig of 786 bp in length
*    32911 33010: gap of      100 bp
*    33011    33754: contig of 744 bp in length
*    33755 33854: gap of      100 bp
*    33855    34575: contig of 721 bp in length
*    34576 34675: gap of      100 bp
*    34676    35415: contig of 740 bp in length
*    35416 35515: gap of      100 bp
*    35516    36256: contig of 741 bp in length
*    36257 36356: gap of      100 bp
*    36357    37094: contig of 738 bp in length
*    37095 37194: gap of      100 bp
*    37195    37899: contig of 705 bp in length
*    37900 37999: gap of      100 bp
*    38000    38751: contig of 752 bp in length
*    38752 38851: gap of      100 bp
*    38852    39616: contig of 765 bp in length
*    39617 39716: gap of      100 bp
*    39717    40458: contig of 742 bp in length
*    40459 40558: gap of      100 bp
*    40559    41317: contig of 759 bp in length
*    41318 41417: gap of      100 bp
*    41418    42187: contig of 770 bp in length
*    42188 42287: gap of      100 bp
*    42288    43028: contig of 741 bp in length
*    43029 43128: gap of      100 bp
*    43129    43842: contig of 714 bp in length
*    43843 43942: gap of      100 bp
*    43943    44668: contig of 726 bp in length
*    44669 44768: gap of      100 bp
*    44769    45555: contig of 787 bp in length
*    45556 45655: gap of      100 bp
*    45656    46397: contig of 742 bp in length
*    46398 46497: gap of      100 bp
*    46498    47273: contig of 776 bp in length
*    47274 47373: gap of      100 bp
*    47374    48135: contig of 762 bp in length
*    48136 48235: gap of      100 bp
*    48236    48972: contig of 737 bp in length
*    48973 49072: gap of      100 bp
*    49073    49818: contig of 746 bp in length
*    49819 49918: gap of      100 bp
*    49919    50698: contig of 780 bp in length.
FEATURES             Location/Qualifiers
     source          1. .50698
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="CTD-2183P16"
                     /clone_lib="CITD Human BAC"
BASE COUNT    14175 a    7874 c    8092 g   14173 t    6384 others
ORIGIN


  Query Match             3.1%;  Score 310;  DB 77;  Length 50698;
  Best Local Similarity   96.3%;  Pred. No. 4.3e-45;
  Matches  316;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     630 tttacaaatatggtccggaaaaagaaccccccctctgagaaacgttgcaagtgaaggcgag 689
           ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19054 TTTGCAGATATGGTCCGGAAAAAGAACCCCCCTCTGAGAAACGTTGCAAGTGAAGGCGAG 18995

Qy     690 ggccagatcctggagcctataggtacagaaagcaaggtatctggaaagaacaaagaattc 749
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18994 GGCCAGATCCTGGAGCCTATAGGTACAGAAAGCAAGGTATCTGGAAAGAACAAAGAATTT 18935

Qy     750 tctgcagatcagatgtcagaaaatacggatcagagtgatgctgcagaactaaatcataag 809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18934 TCTGCAGATCAGATGTCAGAAAATACGGATCAGAGTGATGCTGCAGAACTAAATCATAAG 18875
```

```
Qy     810 gaggaacatagcttgcatgttcaagatccatcttctagcagtaagaaggacttgaaaagc 869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18874 GAGGAACATAGCTTGCATGTTCAAGATCCATCTTCTAGCAGTAAGAAGGACTTGAAAAGC 18815

Qy     870 gcagttctgagtgagaaggctggcttcaattatgaaagccccagtaagggaggaaacttt 929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18814 GCAGTTCTGAGTGAGAAGGCTGGCTTCAATTATGAAAGCCCCAGTAAGGGAGGAAACTTC 18755

Qy     930 ccctcctttccgcatgatgaggtgacag 957
           |  | || |||||||||||||  || |
Db   18754 ACTCCTTTNCCGCATGATGAGGGNACGG 18727


RESULT  14
AC013789/c
LOCUS       AC013789    67407 bp    DNA             HTG       16-JAN-2001
DEFINITION  Homo sapiens clone RP11-19L10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC013789
VERSION     AC013789.5  GI:12232536
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67407)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-19L10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 67407)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 16, 2001 this sequence version replaced gi:11612368.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3853
                Center clone name: 19_L_10
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      719: contig of 719 bp in length
            *      720 819: gap of      100 bp
            *      820     1543: contig of 724 bp in length
            *     1544 1643: gap of      100 bp
            *     1644     2359: contig of 716 bp in length
            *     2360 2459: gap of      100 bp
            *     2460     3180: contig of 721 bp in length
            *     3181 3280: gap of      100 bp
            *     3281     3983: contig of 703 bp in length
            *     3984 4083: gap of      100 bp
            *     4084     4767: contig of 684 bp in length
            *     4768 4867: gap of      100 bp
            *     4868     5577: contig of 710 bp in length
            *     5578 5677: gap of      100 bp
            *     5678     6373: contig of 696 bp in length
            *     6374 6473: gap of      100 bp
            *     6474     7170: contig of 697 bp in length
            *     7171 7270: gap of      100 bp
            *     7271     7963: contig of 693 bp in length
            *     7964 8063: gap of      100 bp
            *     8064     8750: contig of 687 bp in length
            *     8751 8850: gap of      100 bp
            *     8851     9552: contig of 702 bp in length
            *     9553 9652: gap of      100 bp
            *     9653    10372: contig of 720 bp in length
            *    10373 10472: gap of      100 bp
            *    10473    11208: contig of 736 bp in length
            *    11209 11308: gap of      100 bp
            *    11309    12017: contig of 709 bp in length
            *    12018 12117: gap of      100 bp
            *    12118    12809: contig of 692 bp in length
            *    12810 12909: gap of      100 bp
            *    12910    13614: contig of 705 bp in length
            *    13615 13714: gap of      100 bp
            *    13715    14396: contig of 682 bp in length
            *    14397 14496: gap of      100 bp
            *    14497    15167: contig of 671 bp in length
            *    15168 15267: gap of      100 bp
            *    15268    15964: contig of 697 bp in length
            *    15965 16064: gap of      100 bp
            *    16065    16761: contig of 697 bp in length
            *    16762 16861: gap of      100 bp
            *    16862    17598: contig of 737 bp in length
            *    17599 17698: gap of      100 bp
            *    17699    18398: contig of 700 bp in length
            *    18399 18498: gap of      100 bp
            *    18499    19198: contig of 700 bp in length
            *    19199 19298: gap of      100 bp
            *    19299    20014: contig of 716 bp in length
            *    20015 20114: gap of      100 bp
            *    20115    20891: contig of 777 bp in length
            *    20892 20991: gap of      100 bp
            *    20992    21691: contig of 700 bp in length
            *    21692 21791: gap of      100 bp
            *    21792    22485: contig of 694 bp in length
            *    22486 22585: gap of      100 bp
            *    22586    23290: contig of 705 bp in length
            *    23291 23390: gap of      100 bp
            *    23391    24091: contig of 701 bp in length
            *    24092 24191: gap of      100 bp
            *    24192    24885: contig of 694 bp in length
            *    24886 24985: gap of      100 bp
            *    24986    25677: contig of 692 bp in length
            *    25678 25777: gap of      100 bp
            *    25778    26497: contig of 720 bp in length
            *    26498 26597: gap of      100 bp
            *    26598    27295: contig of 698 bp in length
            *    27296 27395: gap of      100 bp
            *    27396    28120: contig of 725 bp in length
            *    28121 28220: gap of      100 bp
            *    28221    28930: contig of 710 bp in length
            *    28931 29030: gap of      100 bp
            *    29031    29715: contig of 685 bp in length
            *    29716 29815: gap of      100 bp
            *    29816    30513: contig of 698 bp in length
            *    30514 30613: gap of      100 bp
```

```
*    30614    31318: contig of 705 bp in length
*    31319 31418: gap of      100 bp
*    31419    32119: contig of 701 bp in length
*    32120 32219: gap of      100 bp
*    32220    32910: contig of 691 bp in length
*    32911 33010: gap of      100 bp
*    33011    33694: contig of 684 bp in length
*    33695 33794: gap of      100 bp
*    33795    34491: contig of 697 bp in length
*    34492 34591: gap of      100 bp
*    34592    35320: contig of 729 bp in length
*    35321 35420: gap of      100 bp
*    35421    36120: contig of 700 bp in length
*    36121 36220: gap of      100 bp
*    36221    36947: contig of 727 bp in length
*    36948 37047: gap of      100 bp
*    37048    37773: contig of 726 bp in length
*    37774 37873: gap of      100 bp
*    37874    38601: contig of 728 bp in length
*    38602 38701: gap of      100 bp
*    38702    39414: contig of 713 bp in length
*    39415 39514: gap of      100 bp
*    39515    40216: contig of 702 bp in length
*    40217 40316: gap of      100 bp
*    40317    41019: contig of 703 bp in length
*    41020 41119: gap of      100 bp
*    41120    41815: contig of 696 bp in length
*    41816 41915: gap of      100 bp
*    41916    42597: contig of 682 bp in length
*    42598 42697: gap of      100 bp
*    42698    43385: contig of 688 bp in length
*    43386 43485: gap of      100 bp
*    43486    44165: contig of 680 bp in length
*    44166 44265: gap of      100 bp
*    44266    44958: contig of 693 bp in length
*    44959 45058: gap of      100 bp
*    45059    45773: contig of 715 bp in length
*    45774 45873: gap of      100 bp
*    45874    46583: contig of 710 bp in length
*    46584 46683: gap of      100 bp
*    46684    47410: contig of 727 bp in length
*    47411 47510: gap of      100 bp
*    47511    48223: contig of 713 bp in length
*    48224 48323: gap of      100 bp
*    48324    49013: contig of 690 bp in length
*    49014 49113: gap of      100 bp
*    49114    49799: contig of 686 bp in length
*    49800 49899: gap of      100 bp
*    49900    50587: contig of 688 bp in length
*    50588 50687: gap of      100 bp
*    50688    51355: contig of 668 bp in length
*    51356 51455: gap of      100 bp
*    51456    52130: contig of 675 bp in length
*    52131 52230: gap of      100 bp
*    52231    52940: contig of 710 bp in length
*    52941 53040: gap of      100 bp
*    53041    53746: contig of 706 bp in length
*    53747 53846: gap of      100 bp
*    53847    54559: contig of 713 bp in length
*    54560 54659: gap of      100 bp
*    54660    55395: contig of 736 bp in length
*    55396 55495: gap of      100 bp
*    55496    56202: contig of 707 bp in length
*    56203 56302: gap of      100 bp
*    56303    57020: contig of 718 bp in length
*    57021 57120: gap of      100 bp
*    57121    57815: contig of 695 bp in length
*    57816 57915: gap of      100 bp
*    57916    58603: contig of 688 bp in length

Query Match              2.6%;  Score 258.4;  DB 63;  Length 67407;
Best Local Similarity   97.8%;  Pred. No. 5.6e-36;
Matches  262;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy   3031 catcctgcgggggagtccgtcatacacccaagcaagcctggggctgctgacgcctgtgtc 3090
           ||||    |||||||||||||||||||||||||||| ||||||||||||||||||||
Db    719 CATCTGCTGGGGGAGTCCGTCATACACCCAAGCAAGCTTGGGGCTGCTGACGCCTGTGTC 660

Qy   3091 tggcacccaagagcagacaaagactctaagggatagtcccaatgtggaggccgcccatct 3150
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    659 TGGCACCCAAGAGCAGACAAAGACTTTAAGGGATAGTCCCAATGTGGAGGCCGCCCATCT 600

Qy   3151 ggcgcgacctatttatggcttggctgtggaaaccaagggattcctgcagggggcgccagc 3210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    599 GGCGCGACCTATTTATGGCTTGGCTGTGGAAACCAAGGGATTCCTGCAGGGGGCGCCAGC 540

Qy   3211 tggcggagagaagtctggggccctcccccagcagtatcctgcatcgggagaaaacaagtc 3270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 TGGCGGAGAGAAGTCTGGGGCCCTCCCCCAGCAGTATCCTGCATCGGGAGAAAACAAGTC 480

Qy   3271 caaggatgaatcccagtccctgttacgg 3298
           |||||||||||||||||||||||||||
Db    479 CAAGGATGAATCCCAGTCCCTGTTACGG 452


RESULT  15
AC013789
LOCUS       AC013789    67407 bp    DNA             HTG       16-JAN-2001
DEFINITION  Homo sapiens clone RP11-19L10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC013789
VERSION     AC013789.5  GI:12232536
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 67407)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-19L10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 67407)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 16, 2001 this sequence version replaced gi:11612368.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3853
                Center clone name: 19_L_10
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
```

```
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      719: contig of 719 bp in length
*      720 819: gap of      100 bp
*      820     1543: contig of 724 bp in length
*     1544 1643: gap of      100 bp
*     1644     2359: contig of 716 bp in length
*     2360 2459: gap of      100 bp
*     2460     3180: contig of 721 bp in length
*     3181 3280: gap of      100 bp
*     3281     3983: contig of 703 bp in length
*     3984 4083: gap of      100 bp
*     4084     4767: contig of 684 bp in length
*     4768 4867: gap of      100 bp
*     4868     5577: contig of 710 bp in length
*     5578 5677: gap of      100 bp
*     5678     6373: contig of 696 bp in length
*     6374 6473: gap of      100 bp
*     6474     7170: contig of 697 bp in length
*     7171 7270: gap of      100 bp
*     7271     7963: contig of 693 bp in length
*     7964 8063: gap of      100 bp
*     8064     8750: contig of 687 bp in length
*     8751 8850: gap of      100 bp
*     8851     9552: contig of 702 bp in length
*     9553 9652: gap of      100 bp
*     9653    10372: contig of 720 bp in length
*    10373 10472: gap of      100 bp
*    10473    11208: contig of 736 bp in length
*    11209 11308: gap of      100 bp
*    11309    12017: contig of 709 bp in length
*    12018 12117: gap of      100 bp
*    12118    12809: contig of 692 bp in length
*    12810 12909: gap of      100 bp
*    12910    13614: contig of 705 bp in length
*    13615 13714: gap of      100 bp
*    13715    14396: contig of 682 bp in length
*    14397 14496: gap of      100 bp
*    14497    15167: contig of 671 bp in length
*    15168 15267: gap of      100 bp
*    15268    15964: contig of 697 bp in length
*    15965 16064: gap of      100 bp
*    16065    16761: contig of 697 bp in length
*    16762 16861: gap of      100 bp
*    16862    17598: contig of 737 bp in length
*    17599 17698: gap of      100 bp
*    17699    18398: contig of 700 bp in length
*    18399 18498: gap of      100 bp
*    18499    19198: contig of 700 bp in length
*    19199 19298: gap of      100 bp
*    19299    20014: contig of 716 bp in length
*    20015 20114: gap of      100 bp
*    20115    20891: contig of 777 bp in length
*    20892 20991: gap of      100 bp
*    20992    21691: contig of 700 bp in length
*    21692 21791: gap of      100 bp
*    21792    22485: contig of 694 bp in length
*    22486 22585: gap of      100 bp
*    22586    23290: contig of 705 bp in length
*    23291 23390: gap of      100 bp
*    23391    24091: contig of 701 bp in length
*    24092 24191: gap of      100 bp
*    24192    24885: contig of 694 bp in length
*    24886 24985: gap of      100 bp
*    24986    25677: contig of 692 bp in length
*    25678 25777: gap of      100 bp
*    25778    26497: contig of 720 bp in length
*    26498 26597: gap of      100 bp
*    26598    27295: contig of 698 bp in length
*    27296 27395: gap of      100 bp
*    27396    28120: contig of 725 bp in length
*    28121 28220: gap of      100 bp
*    28221    28930: contig of 710 bp in length
*    28931 29030: gap of      100 bp
*    29031    29715: contig of 685 bp in length
*    29716 29815: gap of      100 bp
*    29816    30513: contig of 698 bp in length
*    30514 30613: gap of      100 bp
*    30614    31318: contig of 705 bp in length
*    31319 31418: gap of      100 bp
*    31419    32119: contig of 701 bp in length
*    32120 32219: gap of      100 bp
*    32220    32910: contig of 691 bp in length
*    32911 33010: gap of      100 bp
*    33011    33694: contig of 684 bp in length
*    33695 33794: gap of      100 bp
*    33795    34491: contig of 697 bp in length
*    34492 34591: gap of      100 bp
*    34592    35320: contig of 729 bp in length
*    35321 35420: gap of      100 bp
*    35421    36120: contig of 700 bp in length
*    36121 36220: gap of      100 bp
*    36221    36947: contig of 727 bp in length
*    36948 37047: gap of      100 bp
*    37048    37773: contig of 726 bp in length
*    37774 37873: gap of      100 bp
*    37874    38601: contig of 728 bp in length
*    38602 38701: gap of      100 bp
*    38702    39414: contig of 713 bp in length
*    39415 39514: gap of      100 bp
*    39515    40216: contig of 702 bp in length
*    40217 40316: gap of      100 bp
*    40317    41019: contig of 703 bp in length
*    41020 41119: gap of      100 bp
*    41120    41815: contig of 696 bp in length
*    41816 41915: gap of      100 bp
*    41916    42597: contig of 682 bp in length
*    42598 42697: gap of      100 bp
*    42698    43385: contig of 688 bp in length
*    43386 43485: gap of      100 bp
*    43486    44165: contig of 680 bp in length
*    44166 44265: gap of      100 bp
*    44266    44958: contig of 693 bp in length
*    44959 45058: gap of      100 bp
*    45059    45773: contig of 715 bp in length
*    45774 45873: gap of      100 bp
*    45874    46583: contig of 710 bp in length
*    46584 46683: gap of      100 bp
*    46684    47410: contig of 727 bp in length
*    47411 47510: gap of      100 bp
*    47511    48223: contig of 713 bp in length
*    48224 48323: gap of      100 bp
*    48324    49013: contig of 690 bp in length
*    49014 49113: gap of      100 bp
*    49114    49799: contig of 686 bp in length
*    49800 49899: gap of      100 bp
*    49900    50587: contig of 688 bp in length
*    50588 50687: gap of      100 bp
*    50688    51355: contig of 668 bp in length
*    51356 51455: gap of      100 bp
*    51456    52130: contig of 675 bp in length
*    52131 52230: gap of      100 bp
*    52231    52940: contig of 710 bp in length
*    52941 53040: gap of      100 bp
*    53041    53746: contig of 706 bp in length
*    53747 53846: gap of      100 bp
*    53847    54559: contig of 713 bp in length
*    54560 54659: gap of      100 bp
*    54660    55395: contig of 736 bp in length
*    55396 55495: gap of      100 bp
*    55496    56202: contig of 707 bp in length
*    56203 56302: gap of      100 bp
```

```
            *    56303    57020: contig of 718 bp in length
            *    57021 57120: gap of      100 bp
            *    57121    57815: contig of 695 bp in length
            *    57816 57915: gap of      100 bp
            *    57916    58603: contig of 688 bp in length

  Query Match              0.9%;  Score 93.4;  DB 63;  Length 67407;
  Best Local Similarity   98.9%;  Pred. No. 8.7e-07;
  Matches   94;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     547 tcctgctagttgactaatagaattaataattgtaaaaagcactctaaagccacatgcctt 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44306 TCCTGCTAGTTGACTAATAGAATTAATAATTGTAAAAAGCACTCTAAAGCCACATGCCTT 44365

Qy     607 atgaagtcaatgctgggtatgattttacaaatatg 641
           |||||||||||||||||||||||||||| ||||
Db   44366 ATGAAGTCAATGCTGGGTATGATTTTACAAGTATG 44400


Search completed: April 30, 2001, 20:39:51
Job time: 42465 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:07:57 ; Search time 933.78 Seconds
(without alignments)
6255.559 Million cell updates/sec

Title: US-09-721-183-20
Perfect score: 10006
Sequence: 1 ttcctccgcgaaggctcctt.....agttttccaccagttacaca 10006

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9928.2	99.2	10320	21	A54123	Breast cancer prot
2	9925.4	99.2	10319	21	C74759	Human ORFX ORF314
3	995	9.9	1009	21	C79479	cDNA sequence for
4	931.4	9.3	996	22	F44852	Human breast cance
5	668.8	6.7	674	22	F44875	Human breast cance
6	333	3.3	357	21	C16523	Human secreted pro
7	315.4	3.2	333	19	V27248	HeLa gene sequence
8	187	1.9	187	21	C08664	Human secreted pro
9	169.4	1.7	173	21	C21536	Human secreted pro
10	76	0.8	76	21	C28768	Human secreted pro
c 11	72.4	0.7	19124	18	T72882	Plasmodium var-7 g
c 12	72.4	0.7	19124	21	Z98287	Plasmodium var-7 p
c 13	70.6	0.7	6644	20	X33181	Base sequence of t
c 14	70.6	0.7	7372	20	X33182	Base sequence of t
c 15	70.6	0.7	7797	20	X33180	Cowpox virus bsr f
c 16	70.6	0.7	7996	20	X33184	Base sequence of t
c 17	67.2	0.7	9789	17	T41852	cDNA encoding Plas
18	65	0.6	1456	20	Z20844	GATA-3 cDNA sequen
19	65	0.6	2025	21	F21138	Human low adenosin
20	65	0.6	2025	21	A35016	Human adenosine re
c 21	64	0.6	2364	21	A70246	Plasmodium falcipa
c 22	62.2	0.6	5454	21	A70236	Plasmodium falcipa
c 23	61.8	0.6	4590	7	N60472	Sequence encoding
24	61.4	0.6	2688	21	A12375	Murine embryonic b
25	61	0.6	8310	20	Z29911	cDNA encoding a SC
c 26	60.2	0.6	11922	21	A70187	Plasmodium falcipa
27	59.6	0.6	58407	19	V21210	Methanococcus jann
28	58.6	0.6	1984	18	T60490	Mouse Aiolos cDNA.
29	58.6	0.6	1984	20	Z10900	Mouse Aiolos codin
c 30	58.2	0.6	6243	20	Z09598	Clostridium specie
31	57.4	0.6	714	19	V67120	Ikaros protein enc
32	57.4	0.6	1004	17	T16065	Ikaros cDNA. Not
33	57.4	0.6	1004	19	V66974	Ikaros isoform enc
34	57.4	0.6	1004	19	V42811	Human Ikaros cDNA.
35	57.4	0.6	1386	17	T16060	Human Ikaros cDNA
36	57.4	0.6	1386	19	V66969	Human Ikaros encod
37	57.4	0.6	1386	19	V42806	Human Ikaros isofo
38	57.4	0.6	1551	19	V42840	Human Ikaros isofo
39	57.4	0.6	1611	15	Q44980	Human Ikaros pepti
c 40	57.4	0.6	2503	15	Q53480	pNPX30 xylanase cD
41	57.4	0.6	6439	21	C76542	Human ORFX ORF2097
c 42	57.2	0.6	875	21	A01920	Human colon cancer
43	57	0.6	5852	12	Q11710	Dictyostelium plas
44	56.8	0.6	19124	18	T72882	Plasmodium var-7 g
45	56.8	0.6	19124	21	Z98287	Plasmodium var-7 p

ALIGNMENTS

RESULT 1
A54123
ID A54123 standard; DNA; 10320 BP.
XX
AC A54123;
XX
DT 08-FEB-2001 (first entry)
XX
DE Breast cancer protein BCN2 coding sequence.
XX
KW Breast cancer; diagnosis; prognosis; detection; screening;
KW antibody; oestrogen receptor; anti-oestrogen; immune response;
KW lymph node; metastases; tumour; BCR3; BCQ8; BCQ5; BCH1; BCN1; BCN2;
KW BCN5; BCO2; BCX2; BCX3; BCA2; BCR2; BCJ7; BCY3; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 645..4490
FT /*tag= a
FT /product= BCN2 protein
XX
PN WO200055629-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-US06952.
XX
PR 15-MAR-1999; 99US-0268865.
PR 12-NOV-1999; 99US-0439878.
PR 12-NOV-1999; 99US-0440370.
PR 15-NOV-1999; 99US-0440493.
PR 16-NOV-1999; 99US-0440676.
PR 16-NOV-1999; 99US-0440677.

PR 29-NOV-1999; 99US-0450810.
PR 02-DEC-1999; 99US-0453137.
PR 08-MAR-2000; 2000US-0453137.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Mack D, Gish KC;
XX
DR WPI; 2000-638216/61.
DR P-PSDB; B00190.
XX
PT Screening drug candidates for their ability to modulate breast cancer
PT by contacting the drug to a cell expressing an expression profile gene
PT and determining modulation of expression of the gene
XX
PS Disclosure; Fig 46A-D; 258pp; English.
XX
CC New methods for screening drug candidates are described which
CC comprise adding a drug candidate to a cell that expresses a protein
CC selected from BCH1, BCA2, BCJ7, BCN1, BCN5, BCO2, BCQ5, BCR2, BCX2
CC and BCY3 or their fragments and determining the effect of the drug
CC on the expression of those proteins. Antibodies to breast cancer
CC genes (specifically BCH1 or its fragment (BCH1p1 or BCH1p2)) are
CC useful for inhibiting and treating breast cancer in individuals who
CC are non-responsive to anti-oestrogen and positive for oestrogen
CC receptor. Compositions comprising BCH1 or a nucleic acid encoding
CC BCH1 are useful for eliciting an immune response in an individual.
CC The antibodies are also useful for the diagnosis and prognosis of
CC breast cancer and for screening compositions which modulate the
CC breast cancer phenotype. The method allows rapid and simple
CC detection of lymph node metastases.
XX
SQ Sequence 10320 BP; 3174 A; 2005 C; 2096 G; 3042 T; 3 other;

Query Match 99.2%; Score 9928.2; DB 21; Length 10320;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 10001; Conservative 1; Mismatches 4; Indels 11; Gaps 6;

```
Qy       1 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 60

Qy      61 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 120

Qy     121 tctctagcgtcctctttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tctctagcgtcctctttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 180

Qy     181 ctctctcgccttcctatttctttttttttttttttaaacaaaaaacaacacccccctcccct 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ctctctcgccttcctatttctttttttttttttttaaacaaaaaacaacacccccctcccct 240

Qy     241 ctcccacccggcaccgggcacatccttgctctatttcctttctctt----tctctctc 294
          |||||||||||||||||||||||||||||||||||||||||||||      ||||||||
Db     241 ctcccacccggcaccgggcacatccttgctctatttcctttctctttctctctctctctc 300

Qy     295 tctctctctctttttttaataagggtgggggagggaaaggggggggaggcaggaaagacct 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tctctctctctttttttaataagggtgggggagggaaaggggggggatgcaggaaagacct 360

Qy     355 ttttctctcccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcat 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ttttctctcccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcat 420

Qy     415 ggctttggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatca 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ggctttggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatca 480

Qy     475 acagatgtacagatcagctctcaaaatgtcttctgtgtcttctgagcgtcttctaagaca 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 acagatgtacagatcagctctcaaaatgtcttctgtgtcttctgagcgtcttctaagaca 540

Qy     535 attgcattagcctcctgctagttgactaatagaattaataattgtaaaaagcactctaaa 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 attgcattagcctcctgctagttgactaatagaattaataattgtaaaaagcactctaaa 600

Qy     595 gccacatgccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaaga 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 gccacatgccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaaga 660

Qy     655 acccccctctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggta 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 acccccctctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggta 720

Qy     715 cagaaagcaaggtatctggaaagaacaaagaattctctgcagatcagatgtcagaaaata 774
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     721 cagaaagcaaggtatctggaaagaacaaagaattttctgcagatcagatgtcagaaaata 780

Qy     775 cggatcagagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaag 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 cggatcagagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaag 840

Qy     835 atccatcttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggct 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 atccatcttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggct 900

Qy     895 tcaattatgaaagccccagtaagggaggaaactttccctcctttccgcatgatgaggtga 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 tcaattatgaaagccccagtaagggaggaaactttccctcctttccgcatgatgaggtga 960

Qy     955 cagacagaaatatgttggctttctcatttccagctgctgggggagtctgtgagcccttga 1014
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     961 cagacagaaatatgttggctttctcatctccagctgctgggggagtctgtgagcccttga 1020

Qy    1015 agtctccgcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcagggg 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 agtctccgcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcagggg 1080

Qy    1075 actcactggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggc 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 actcactggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggc 1140

Qy    1135 aagcacagagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaa 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 aagcacagagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaa 1200

Qy    1195 acccacaagtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactgg 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 acccacaagtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactgg 1260

Qy    1255 tgaatgacaacccagacccggcacctctgtctccagagcttcaggactttaaatgcaata 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 tgaatgacaacccagacccggcacctctgtctccagagcttcaggactttaaatgcaata 1320

Qy    1315 tctgtggatatggttactacggcaacgaccccacagatctgattaagcacttccgaaagt 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 tctgtggatatggttactacggcaacgaccccacagatctgattaagcacttccgaaagt 1380

Qy    1375 atcacttaggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttgg 1434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 atcacttaggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttgg 1440

Qy    1435 cccttcataacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctg 1494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 cccttcataacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctg 1500

Qy    1495 tgttttctggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacct 1554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 tgttttctggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacct 1560

Qy    1555 atgatgtgcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagatt 1614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 atgatgtgcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagatt 1620
```

```
Qy   1615 gccaagggaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaact 1674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 gccaagggaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaact 1680

Qy   1675 catccaccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctc 1734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 catccaccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctc 1740

Qy   1735 tcccctcctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcac 1794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 tcccctcctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcac 1800

Qy   1795 ttcaatccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcag 1854
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ttcaatccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcag 1860

Qy   1855 gagatgacactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaa 1914
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gagatgacactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaa 1920

Qy   1915 a-ggtacagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcat 1973
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 atggtacagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcat 1980

Qy   1974 ctagctcacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcg 2033
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ctagctcacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcg 2040

Qy   2034 gccttaatccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatc 2093
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 gccttaatccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatc 2100

Qy   2094 tagccaaaagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggcta 2153
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 tagccaaaagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggcta 2160

Qy   2154 aaaagaaggacttctccagcaagggagccgaggataatatggtaacgagctataattgtc 2213
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 aaaagaaggacttctccagcaagggagccgaggataatatggtaacgagctataattgtc 2220

Qy   2214 agttctgtgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccac 2273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 agttctgtgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccac 2280

Qy   2274 ttctccgtcattatcaacagctccataacattcacaagtgtaccattaaacactgtccat 2333
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 ttctccgtcattatcaacagctccataacattcacaagtgtaccattaaacactgtccat 2340

Qy   2334 tctgtcccagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttg 2393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 tctgtcccagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttg 2400

Qy   2394 cttgtagaaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctgggg 2453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 cttgtagaaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctgggg 2460

Qy   2454 cggctggaagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacg 2513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 cggctggaagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacg 2520

Qy   2514 tagatgtactcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaac 2573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 tagatgtactcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaac 2580

Qy   2574 aagaagcaaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaac 2633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 aagaagcaaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaac 2640

Qy   2634 actcatgtaccaaatgtgattttattacccaagtggaagaagagatttcccgacactaca 2693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 actcatgtaccaaatgtgattttattacccaagtggaagaagagatttcccgacactaca 2700
```

```
Qy   2694 ggagagcacacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagt 2753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 ggagagcacacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagt 2760

Qy   2754 cactactggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacg 2813
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 cactactggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacg 2820

Qy   2814 gcgaagaggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcaggg 2873
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 gcgaagaggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcaggg 2880

Qy   2874 tctacaatctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtga 2933
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 tctacaatctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtga 2940

Qy   2934 agagagagaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttcca 2993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 agagagagaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttcca 3000

Qy   2994 gtgatgaccttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcat 3053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 gtgatgaccttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcat 3060

Qy   3054 acacccaagcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaaga 3113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 acacccaagcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaaga 3120

Qy   3114 ctctaagggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttgg 3173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 ctctaagggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttgg 3180

Qy   3174 ctgtggaaaccaagggattcctgcaggggcgccagctggcggagagaagtctggggccc 3233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 ctgtggaaaccaagggattcctgcaggggcgccagctggcggagagaagtctggggccc 3240

Qy   3234 tcccccagcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgt 3293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 tcccccagcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgt 3300

Qy   3294 tacggaggcgtagaggctccggtgtttttttgtgccaattgcctgaccacaaagacctctc 3353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 tacggaggcgtagaggctccggtgtttttttgtgccaattgcctgaccacaaagacctctc 3360

Qy   3354 tctggcgaaagaatgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagc 3413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 tctggcgaaagaatgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagc 3420

Qy   3414 ttcactcgactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattatta 3473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 ttcactcgactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattatta 3480

Qy   3474 ggaggagaacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaac 3533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 ggaggagaacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaac 3540

Qy   3534 agcagaggggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaag 3593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 agcagaggggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaag 3600

Qy   3594 atcatctaactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaag 3653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 atcatctaactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaag 3660

Qy   3654 cccagggttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactc 3713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 cccagggttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactt 3720

Qy   3714 tggatattcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagta 3773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 tggatattcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagta 3780

Qy   3774 ctggagatccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggca 3833
```

```
Db  3781 ctggagatccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggca 3840
Qy  3834 gtcctatagaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagcc 3893
Db  3841 gtcctatagaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagcc 3900
Qy  3894 ctattgaaaagtaccagtacccacttttggacttcccttt-tacataatgacttccaga 3952
Db  3901 ctattgaaaagtaccagtacccacttttggacttcccttgtacataatgacttccaga 3960
Qy  3953 gtgaagctgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgc 4012
Db  3961 gtgaagctgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgc 4020
Qy  4013 actacttgagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatccca 4072
Db  4021 actacttgagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatccca 4080
Qy  4073 ccttcaatctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatt 4132
Db  4081 ccttcaatctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatt 4140
Qy  4133 tggcgatcaagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacga 4192
Db  4141 tggcgatcaagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacga 4200
Qy  4193 aggcaccaccaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttg 4252
Db  4201 aggcaccaccaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttg 4260
Qy  4253 atagaagtactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctgg 4312
Db  4261 atagaagtactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctgg 4320
Qy  4313 atgaagtgatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgca 4372
Db  4321 atgaagtgatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgca 4380
Qy  4373 gcatatgccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcc 4432
Db  4381 gcatatgccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcc 4440
Qy  4433 tgcataggaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagca 4492
Db  4441 tgcataggaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagca 4500
Qy  4493 cttagcacaattaaatagaaatagggttttcttgatgggaattcaatagcttgtaatgtct 4552
Db  4501 cttagcacaattaaatagaaatagggttttcttgatgggaattcaatagcttgtaatgtct 4560
Qy  4553 tatgaagacctattaaaaaaatacttcatagagcctgccttatccaacatgaaattccct 4612
Db  4561 tatgaagacctattaaaaaaatacttcatagagcctgccttatccaacatgaaattccct 4620
Qy  4613 tcttttgttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatgg 4672
Db  4621 tcttttgttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatgg 4680
Qy  4673 tcaatgagaaagaatggaagatggtaaacaatcacttttttaaaacctgttaagtcaaaac 4732
Db  4681 tcaatgagaaagaatggaagatggtaaacaatcacttttttaaaacctgttaagtcaaaac 4740
Qy  4733 catcttggctaatatgtactggggaaataatccataagagatatcaccagactagaatta 4792
Db  4741 catcttggctaatatgtactggggaaataatccataagagatatcaccagactagaatta 4800
Qy  4793 atatatttataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattatt 4852
Db  4801 atatatttataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattatt 4860
Qy  4853 atactaaagcagtactggactggccattggaccatttgttccaaaacccataaattgttg 4912
Db  4861 atactaaagcagtactggactggccattggaccatttgttccaaaacccataaattgttg 4920
Qy  4913 cctaaatttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggc 4972
Db  4921 cctaaatttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggc 4980
Qy  4973 aatgaaagaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagatt 5032
Db  4981 aatgaaagaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagatt 5040
Qy  5033 ttgacctagaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacag 5092
Db  5041 ttgacctagaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacag 5100
Qy  5093 atggtggcacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctga 5152
Db  5101 atggtggcacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctga 5160
Qy  5153 tatttttgtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattca 5212
Db  5161 tatttttgtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattca 5220
Qy  5213 ttttgtccacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgattt 5272
Db  5221 ttttgtccacacatgtacccatttttaaaaaacaatgtcctcgatgcttctgtagtgattt 5280
Qy  5273 cattttagccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaa 5332
Db  5281 cattttagccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaa 5340
Qy  5333 gcctcctgttggttactaatctcacttggcacattataactaaaggaatcccctcaattc 5392
Db  5341 gcctcctgttggttactaatctcacttggcacattataactaaaggaatcccctcaattc 5400
Qy  5393 aaaagcatagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggca 5452
Db  5401 aaaagcatagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggca 5460
Qy  5453 tttcttcacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttcca 5512
Db  5461 tttcttcacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttcca 5520
Qy  5513 aactttgtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgt 5572
Db  5521 aactttgtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgt 5580
Qy  5573 accagccaatcccttcattaaatgtatacagatttagttaagtagcattaaataggattc 5632
Db  5581 accagccaatcccttcattaaatgtatacagatttagttaagtagcattaaataggattc 5640
Qy  5633 ttagaagtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaa 5692
Db  5641 ttagaagtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaa 5700
Qy  5693 gggaaagctcctcagcataactgctcagggaaatagggctaaataactgaacattaaata 5752
Db  5701 gggaaagctcctcagcataactgctcagggaaatagggctaaataactgaacattaaata 5760
Qy  5753 attggttaaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaagg 5812
Db  5761 attggttaaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaagg 5820
Qy  5813 actgactttaataatttgcattatattgtcccaaccagtagtttatttttgccacggag 5872
Db  5821 actgactttaataatttgcattatattgtcccaaccagtagtttatttttgccacggag 5880
Qy  5873 atgtagaagatattacaagctactggatgcactgtcagattaacttatttcattaaagaa 5932
Db  5881 atgtagaagatattacaagctactggatgcactgtcagattaacttatttcattaaagaa 5940
Qy  5933 gttgggagaacaaatagg-aaaaaaaaaacttatttttctagtaaatattaatgtattaca 5991
Db  5941 gttgggagaacaaatagg-aaaaaaaaaacttatttttctagtaaatattaatgtattaca 6000
```

```
Qy   5992 tttcaaataatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaaca 6051
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 tttcaaataatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaaca 6060

Qy   6052 aagatattccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaa 6111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 aagatattccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaa 6120

Qy   6112 atgtagcaaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgt 6171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6121 atgtagcaaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgt 6180

Qy   6172 gtgtgtgtgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataat 6231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6181 gtgtgtgtgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataat 6240

Qy   6232 ggtatggaaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggta 6291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6241 ggtatggaaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggta 6300

Qy   6292 gcacttattcattccaagtacttttttttttttaatttttaagctcttaactcacattgt 6351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6301 gcacttattcattccaagtacttttttttttttaatttttaagctcttaactcacattgt 6360

Qy   6352 tatgcttaagatgataaacatatatcctcttttattgctttgtctatgtttcatatgaa 6411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6361 tatgcttaagatgataaacatatatcctcttttattgctttgtctatgtttcatatgaa 6420

Qy   6412 acatttcagaaattattttgataagtgttgctggaatctgcaacgctgatttttttttgc 6471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6421 acatttcagaaattattttgataagtgttgctggaatctgcaacgctgatttttttttgc 6480

Qy   6472 attctgtagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattg 6531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6481 attctgtagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattg 6540

Qy   6532 ttttgtttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatg 6591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6541 ttttgtttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatg 6600

Qy   6592 gcaggtagagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtcttt 6651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6601 gcaggtagagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtcttt 6660

Qy   6652 ctgatgttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagaga 6711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6661 ctgatgttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagaga 6720

Qy   6712 aaagagaaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaac 6771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6721 aaagagaaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaac 6780

Qy   6772 gggccttattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaag 6831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6781 gggccttattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaag 6840

Qy   6832 aagtggaatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaat 6891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6841 aagtggaatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaat 6900

Qy   6892 tttgatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaa 6951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6901 tttgatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaa 6960

Qy   6952 aaaagagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacag 7011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6961 aaaagagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacag 7020

Qy   7012 aatgcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaac 7071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7021 aatgcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaac 7080

Qy   7072 accagaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactc 7131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7081 accagaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactc 7140

Qy   7132 attaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgc 7191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7141 attaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgc 7200

Qy   7192 ctttttatttttcattcaatttttttttttttctcagacagacatagtagtatcaactagca 7251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7201 ctttttatttttcattcaatttttttttttttctcagacagacatagtagtatcaactagca 7260

Qy   7252 ttggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaat 7311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7261 ttggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaat 7320

Qy   7312 atttttggatagcgttgacacgatagatcttattccatacttctttattattgataattt 7371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7321 atttttggatagcgttgacacgatagatcttattccatacttctttattattgataattt 7380

Qy   7372 tattttcatttttgctttcattattatacatattttggtggagaagaggttgggctttt 7431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7381 tattttcatttttgctttcattattatacatattttggtggagaagaggttgggctttt 7440

Qy   7432 ttgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagcct 7491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7441 ttgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagcct 7500

Qy   7492 ttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatt 7551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7501 ttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatt 7560

Qy   7552 attttatcttaaattgtgattttaaacgagctattatggtacggaacttttttaatgag 7611
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7561 attttatcttaaattgtgattttaaacgagctattatggtacggaacttttttaatgag 7620

Qy   7612 gaatttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaat 7671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7621 gaatttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaat 7680

Qy   7672 gatgtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtc 7731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7681 gatgtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtc 7740

Qy   7732 tatattctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcat 7791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7741 tatattctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcat 7800

Qy   7792 tcttcctttttctcactctttctctcatcactcacctctgtattcattctgttgtttggga 7851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7801 tcttcctttttctcactctttctctcatcactcacctctgtattcattctgttgtttggga 7860

Qy   7852 tagaaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactaca 7911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7861 tagaaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactaca 7920

Qy   7912 tgactccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg 7971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7921 tgactccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg 7980

Qy   7972 -aaagagtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttа 8030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7981 aaaagagtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttа 8040

Qy   8031 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaaca 8090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8041 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaaca 8100

Qy   8091 agcatggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacag 8150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8101 agcatggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacag 8160

Qy   8151 ctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatct 8210
```

```
Db   8161 ctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatct 8220

Qy   8211 attatgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgc 8270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8221 attatgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgc 8280

Qy   8271 ctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataatt 8330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8281 ctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataatt 8340

Qy   8331 attagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtga 8390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8341 attagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtga 8400

Qy   8391 actgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatc 8450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8401 actgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatc 8460

Qy   8451 cagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcct 8510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8461 cagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcct 8520

Qy   8511 tttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 8570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8521 tttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 8580

Qy   8571 aaaaaaaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagctt 8630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8581 aaaaaaaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagctt 8640

Qy   8631 gctgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaac 8690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8641 gctgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaac 8700

Qy   8691 taaatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaa 8750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8701 taaatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaa 8760

Qy   8751 atgagaagaagatagaaggacgcccgtcgggactttggagggcattgttattttcccaaa 8810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8761 atgagaagaagatagaaggacgcccgtcgggactttggagggcattgttattttcccaaa 8820

Qy   8811 gaaagacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttca 8870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8821 gaaagacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttca 8880

Qy   8871 gtactgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaag 8930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8881 gtactgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaag 8940

Qy   8931 catcccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaa 8990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8941 catcccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaa 9000

Qy   8991 gtatgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaa 9050
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9001 gtatgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaa 9060

Qy   9051 tgtacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctac 9110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9061 tgtacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctac 9120

Qy   9111 accggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaa 9170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9121 accggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaa 9180

Qy   9171 attgtccttctagagcatagagtcccaaaactcaattctggttttcccctgttttttttt 9230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9181 attgtccttctagagcatagagtcccaaaactcaattctggttttcccctgttttttttt 9240

Qy   9231 ttttttttttttcccaacatatgaactgcagcatatcacttttttctttttgtgcctcaggt 9290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   9241 ttttttttttttcccaacatatgaactgcagcatatcacttttttctttttgtgcctcaggt 9300

Qy   9291 tcctcacctgtaaaattgaaaaatatatgtattaataatattattaataataataatggt 9350
          |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9301 tcctcasctgtaaaattgaaaaatatatgtattaataatattattaataataataatggt 9360

Qy   9351 aatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtg 9410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9361 aatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtg 9420

Qy   9411 ccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaa 9470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9421 ccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaa 9480

Qy   9471 aggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgt 9530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9481 aggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgt 9540

Qy   9531 tgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctac 9590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9541 tgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctac 9600

Qy   9591 caataaggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaac 9650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9601 caataaggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaac 9660

Qy   9651 cacatggaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaat 9710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9661 cacatggaaaactttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaat 9720

Qy   9711 ttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttg 9770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9721 ttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttg 9780

Qy   9771 ctaagaaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatag 9830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9781 ctaagaaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatag 9840

Qy   9831 ttctgttcttacaacacatgaaatttttttcgttttatttttatttttgttttcatagtgcat 9890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9841 ttctgttcttacaacacatgaaatttttttcgttttatttttatttttgttttcatagtgcat 9900

Qy   9891 gttcatttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggca 9950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9901 gttcatttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggca 9960

Qy   9951 gcaaagatgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   9961 gcaaagatgtctgttacatctaaacttgaataataaagttttaccaccagttacaca 10017


RESULT   2
C74759
ID   C74759 standard; cDNA; 10319 BP.
XX
AC   C74759;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF314 polynucleotide sequence SEQ ID NO:627.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
```

```
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; B40550.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 743-750; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 10319 BP; 3176 A; 2006 C; 2093 G; 3043 T; 1 other;

  Query Match             99.2%;  Score 9925.4;  DB 21;  Length 10319;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 10000;  Conservative    0;  Mismatches    6;  Indels   11;  Gaps    6;

Qy       1 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 ttcctccgcgaaggctcctttgatattaatagtgttggtgtcttgaaactgacgtaatgc 61

Qy      61 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 gcggagactgaggtcctgacaagcgataacatttctgataaagacccgatcttactgcaa 121

Qy     121 tctctagcgtcctctttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 tctctagcgtcctctttttttggtgctgctggtttctccagacctcgcgtcctctcgattg 181

Qy     181 ctctctcgccttcctatttctttttttttttttttaaacaaaaaacaacacccctcccct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 ctctctcgccttcctatttctttttttttttttttaaacaaaaaacaacacccctcccct 241

Qy     241 ctcccacccggcaccgggcacatccttgctctatttccttttctctt------tctctctc 294
           |||||||||||||||||||||||||||||||||||||||||||||      ||||||||
Db     242 ctcccacccggcaccgggcacatccttgctctatttccttttctctttctctctctctctc 301

Qy     295 tctctctctctttttttaataagggtgggggagggaaaggggggggaggcaggaaagacct 354
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     302 tctctctctctttttttaataagggtgggggagggaaaggggggggatgcaggaaagacct 361

Qy     355 ttttctctccccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcat 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 ttttctctccccccccgcaataatccaagatcaactctgcaaacaacagaagacggttcat 421

Qy     415 ggctttggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatca 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 ggctttggccgccgcgccaccatctttcgggctgccgagggtgttcttgacgattaatca 481

Qy     475 acagatgtacagatcagctctcaaaatgtcttctgtgtcttctgagcgtcttctaagaca 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 acagatgtacagatcagctctcaaaatgtcttctgtgtcttctgagcgtcttctaagaca 541

Qy     535 attgcattagcctcctgctagttgactaatagaattaataattgtaaaaagcactctaaa 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 attgcattagcctcctgctagttgactaatagaattaataattgtaaaaagcactctaaa 601

Qy     595 gccacatgccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaaga 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 gccacatgccttatgaagtcaatgctgggtatgattttacaaatatggtccggaaaaaga 661

Qy     655 acccccctctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggta 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 acccccctctgagaaacgttgcaagtgaaggcgagggccagatcctggagcctataggta 721

Qy     715 cagaaagcaaggtatctggaaagaacaaagaattctctgcagatcagatgtcagaaaata 774
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     722 cagaaagcaaggtatctggaaagaacaaagaatttttctgcagatcagatgtcagaaaata 781

Qy     775 cggatcagagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaag 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 cggatcagagtgatgctgcagaactaaatcataaggaggaacatagcttgcatgttcaag 841

Qy     835 atccatcttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggct 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 atccatcttctagcagtaagaaggacttgaaaagcgcagttctgagtgagaaggctggct 901

Qy     895 tcaattatgaaagccccagtaagggaggaaactttccctccttccgcatgatgaggtga 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 tcaattatgaaagccccagtaagggaggaaactttccctccttccgcatgatgaggtga 961

Qy     955 cagacagaaatatgttggctttctcatttccagctgctgggggagtctgtgagcccttga 1014
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     962 cagacagaaatatgttggctttctcatctccagctgctgggggagtctgtgagcccttga 1021

Qy    1015 agtctccgcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcagggg 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1022 agtctccgcaaagagcagaggcagatgaccctcaagatatggcctgcacccccctcagggg 1081

Qy    1075 actcactggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggc 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1082 actcactggagacaaaggaagatcagaagatgtcaccaaaggctacagaggaaacagggc 1141

Qy    1135 aagcacagagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaa 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1142 aagcacagagtggtcaagccaattgtcaaggtttgagcccagtttcagtggcctcaaaaa 1201

Qy    1195 acccacaagtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactgg 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1202 acccacaagtgccttcagatgggggtgtaagactgaataaatccaaaactgacttactgg 1261

Qy    1255 tgaatgacaacccagacccggcacctctgtctccagagcttcaggactttaaatgcaata 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1262 tgaatgacaacccagacccggcacctctgtctccagagcttcaggactttaaatgcaata 1321

Qy    1315 tctgtggatatggttactacggcaacgaccccacagatctgattaagcacttccgaaagt 1374
```

```
Db  1322 tctgtggatatggttactacggcaacgaccccacagatctgattaagcacttccgaaagt 1381
Qy  1375 atcacttaggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttgg 1434
Db  1382 atcacttaggactgcataaccgcaccaggcaagatgctgagctggacagcaaaatcttgg 1441
Qy  1435 cccttcataacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctg 1494
Db  1442 cccttcataacatggtgcagttcagccattccaaagacttccagaaggtcaaccgttctg 1501
Qy  1495 tgttttctggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacct 1554
Db  1502 tgttttctggtgtgctgcaggacatcaattcttcaaggcctgttttactaaatgggacct 1561
Qy  1555 atgatgtgcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagatt 1614
Db  1562 atgatgtgcaggtgacttcaggtggaacattcattggcattggacggaaaacaccagatt 1621
Qy  1615 gccaagggaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaact 1674
Db  1622 gccaagggaacaccaagtatttccgctgtaaattctgcaatttcacttatatgggcaact 1681
Qy  1675 catccaccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctc 1734
Db  1682 catccaccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctc 1741
Qy  1735 tcccctcctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcac 1794
Db  1742 tcccctcctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcac 1801
Qy  1795 ttcaatccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcag 1854
Db  1802 ttcaatccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcag 1861
Qy  1855 gagatgacactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaa 1914
Db  1862 gagatgacactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaa 1921
Qy  1915 a-ggtacagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcat 1973
Db  1922 atggtacagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcat 1981
Qy  1974 ctagctcacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcg 2033
Db  1982 ctagctcacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcg 2041
Qy  2034 gccttaatccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatc 2093
Db  2042 gccttaatccagagttaaatgataagctttccaggggctctgtcattaatcagaatgatc 2101
Qy  2094 tagccaaaagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggcta 2153
Db  2102 tagccaaaagttcagaaggagagacaatgaccaagacagacaagagctcgagtggggcta 2161
Qy  2154 aaaagaaggacttctccagcaagggagccgaggataatatggtaacgagctataattgtc 2213
Db  2162 aaaagaaggacttctccagcaagggagccgaggataatatggtaacgagctataattgtc 2221
Qy  2214 agttctgtgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccac 2273
Db  2222 agttctgtgacttccgatattccaaaagccatggccctgatgtaattgtagtggggccac 2281
Qy  2274 ttctccgtcattatcaacagctccataacattcacaagtgtaccattaaacactgtccat 2333
Db  2282 ttctccgtcattatcaacagctccataacattcacaagtgtaccattaaacactgtccat 2341
Qy  2334 tctgtcccagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttg 2393
Db  2342 tctgtcccagaggactttgcagcccagaaaagcaccttggagaaattacttatccgtttg 2401
Qy  2394 cttgtagaaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctgggg 2453
Db  2402 cttgtagaaaaagtaattgttcccactgtgcactcttgcttctgcacttgtctcctgggg 2461
Qy  2454 cggctggaagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacg 2513
Db  2462 cggctggaagctcgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacg 2521
Qy  2514 tagatgtactcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaac 2573
Db  2522 tagatgtactcctctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaac 2581
Qy  2574 aagaagcaaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaac 2633
Db  2582 aagaagcaaatcacctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaac 2641
Qy  2634 actcatgtaccaaatgtgattttattacccaagtggaagaagagatttcccgacactaca 2693
Db  2642 actcatgtaccaaatgtgattttattacccaagtggaagaagagatttcccgacactaca 2701
Qy  2694 ggagagcacacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagt 2753
Db  2702 ggagagcacacagctgctacaaatgccgtcagtgcagttttacagctgccgatactcagt 2761
Qy  2754 cactactggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacg 2813
Db  2762 cactactggagcacttcaacactgttcactgccaggaacaggacatcactacagccaacg 2821
Qy  2814 gcgaagaggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcaggg 2873
Db  2822 gcgaagaggacggtcatgccatatccaccatcaaagaggagcccaaaattgacttcaggg 2881
Qy  2874 tctacaatctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtga 2933
Db  2882 tctacaatctgctaactccagactctaaaatgggagagccagtttctgagagtgtggtga 2941
Qy  2934 agagagagaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttcca 2993
Db  2942 agagagagaagctggaagagaaggacgggctcaaagagaaagtttggaccgagagttcca 3001
Qy  2994 gtgatgaccttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcat 3053
Db  3002 gtgatgaccttcgcaatgtgacttggagaggggcagacatcctgcgggggagtccgtcat 3061
Qy  3054 acacccaagcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaaga 3113
Db  3062 acacccaagcaagcctggggctgctgacgcctgtgtctggcacccaagagcagacaaaga 3121
Qy  3114 ctctaagggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttgg 3173
Db  3122 ctctaagggatagtcccaatgtggaggccgcccatctggcgcgacctatttatggcttgg 3181
Qy  3174 ctgtggaaaccaagggattcctgcaggggcgccagctggcggagagaagtctggggccc 3233
Db  3182 ctgtggaaaccaagggattcctgcaggggcgccagctggcggagagaagtctggggccc 3241
Qy  3234 tcccccagcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgt 3293
Db  3242 tcccccagcagtatcctgcatcgggagaaaacaagtccaaggatgaatcccagtccctgt 3301
Qy  3294 tacggaggcgtagaggctccggtgtttttgtgccaattgcctgaccacaaagacctctc 3353
Db  3302 tacggaggcgtagaggctccggtgtttttgtgccaattgcctgaccacaaagacctctc 3361
Qy  3354 tctggcgaaagaatgcaaatggcggatatgtatgcaacgcgtgtggcctctaccagaagc 3413
Db  3362 tctggcgaaagaatgcaaatggcggatatgtatgcaacgcgtatggcctctaccagaagc 3421
Qy  3414 ttcactcgactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattatta 3473
Db  3422 ttcactcgactcccaggcctttaaacatcattaaacaaaacaacggtgagcagattatta 3481
Qy  3474 ggaggagaacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaac 3533
Db  3482 ggaggagaacaagaaagcgccttaacccagaggcacttcaggctgagcagctcaacaaac 3541
```

```
Qy   3534 agcagaggggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaag 3593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3542 agcagaggggcagcaatgaggagcaagtcaatggaagcccgttagagaggaggtcagaag 3601

Qy   3594 atcatctaactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaag 3653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3602 atcatctaactgaaagtcaccagagagaaattccactccccagcctaagtaaatacgaag 3661

Qy   3654 cccagggttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactc 3713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3662 cccagggttcattgactaaaagccattctgctcagcagccagtcctggtcagccaaactc 3721

Qy   3714 tggatattcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagta 3773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3722 tggatattcacaaaaggatgcaacctttgcacattcagataaaaagtcctcaggaaagta 3781

Qy   3774 ctggagatccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggca 3833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3782 ctggagatccaggaaatagttcatccgtatctgaagggaaaggaagttctgagagaggca 3841

Qy   3834 gtcctatagaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagcc 3893
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3842 gtcctatagaaaagtacatgagacctgcgaaacacccaaattattcaccaccaggcagcc 3901

Qy   3894 ctattgaaaagtaccagtacccacttttggacttcccttt-tacataatgacttccaga 3952
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   3902 ctattgaaaagtaccagtacccacttttggacttcccttgtacataatgacttccaga 3961

Qy   3953 gtgaagctgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgc 4012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3962 gtgaagctgattggctgcggttctggagtaaatataagctctccgttcctgggaatccgc 4021

Qy   4013 actacttgagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatccca 4072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4022 actacttgagtcacgtgcctggcctaccaaatccttgccaaaactatgtgccttatccca 4081

Qy   4073 ccttcaatctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatt 4132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4082 ccttcaatctgcctcctcatttttcagctgttggatcagacaatgacattcctctagatt 4141

Qy   4133 tggcgatcaagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacga 4192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4142 tggcgatcaagcattccagacctgggccaactgcaaacggtgcctccaaggagaaaacga 4201

Qy   4193 aggcaccaccaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttg 4252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4202 aggcaccaccaaatgtaaaaaatgaaggtcccttgaatgtagtaaaaacagagaaagttg 4261

Qy   4253 atagaagtactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctgg 4312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4262 atagaagtactcaagatgaactttcaacaaaatgtgtgcactgtggcattgtctttctgg 4321

Qy   4313 atgaagtgatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgca 4372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4322 atgaagtgatgtatgctttgcatatgagttgccatggtgacagtggacctttccagtgca 4381

Qy   4373 gcatatgccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcc 4432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4382 gcatatgccagcatctttgcacggacaaatatgacttcacaacacatatccagaggggcc 4441

Qy   4433 tgcataggaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagca 4492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4442 tgcataggaacaatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagca 4501

Qy   4493 cttagcacaattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtct 4552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4502 cttagcacaattaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtct 4561

Qy   4553 tatgaagacctattaaaaaaatacttcatagagcctgccttatccaacatgaaattccct 4612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4562 tatgaagacctattaaaaaaatacttcatagagcctgccttatccaacatgaaattccct 4621

Qy   4613 tcttttgttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatgg 4672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4622 tcttttgttattctttcttttgatgagtaggttaccaagattaaaaagtgagataaatgg 4681

Qy   4673 tcaatgagaaagaatggaagatggtaaacaatcacttttaaaacctgttaagtcaaaac 4732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4682 tcaatgagaaagaatggaagatggtaaacaatcacttttaaaacctgttaagtcaaaac 4741

Qy   4733 catcttggctaatatgtactggggaaataatccataagagatatcaccagactagaatta 4792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4742 catcttggctaatatgtactggggaaataatccataagagatatcaccagactagaatta 4801

Qy   4793 atatatttataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattatt 4852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4802 atatatttataaagaaagagaccaaaactgtctagaatttgaaagggtttacatattatt 4861

Qy   4853 atactaaagcagtactggactggccattggaccatttgttccaaaacccataaattgttg 4912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4862 atactaaagcagtactggactggccattggaccatttgttccaaaacccataaattgttg 4921

Qy   4913 cctaaatttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggc 4972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4922 cctaaatttataatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggc 4981

Qy   4973 aatgaaagaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagatt 5032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4982 aatgaaagaaaaatggcgccctctcaatttagtcttctctcattggccatgtttcagatt 5041

Qy   5033 ttgacctagaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacag 5092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5042 ttgacctagaaatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacag 5101

Qy   5093 atggtggcacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctga 5152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5102 atggtggcacgctgtttttacccagccctgcctgtacatacacatgcacaccctctctga 5161

Qy   5153 tatttttgtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattca 5212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5162 tatttttgtcctttagatgttcaaatactcagtagtccttttgtttgcggtttagattca 5221

Qy   5213 ttttgtccacacatgtacccatttaaaaaacaatgtcctcgatgcttctgtagtgattt 5272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5222 ttttgtccacacatgtacccatttaaaaaacaatgtcctcgatgcttctgtagtgattt 5281

Qy   5273 cattttagccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaa 5332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5282 cattttagccaggtatttctttcttgtgtgtgatgaaccagtatggatttgcttttctaa 5341

Qy   5333 gcctcctgttggttactaatctcacttggcacattataactaaaggaatcccctcaattc 5392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5342 gcctcctgttggttactaatctcacttggcacattataactaaaggaatcccctcaattc 5401

Qy   5393 aaaagcatagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggca 5452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5402 aaaagcatagatggatacaaatgtcagaccgtgggtttaatttgtttagaacacatggca 5461

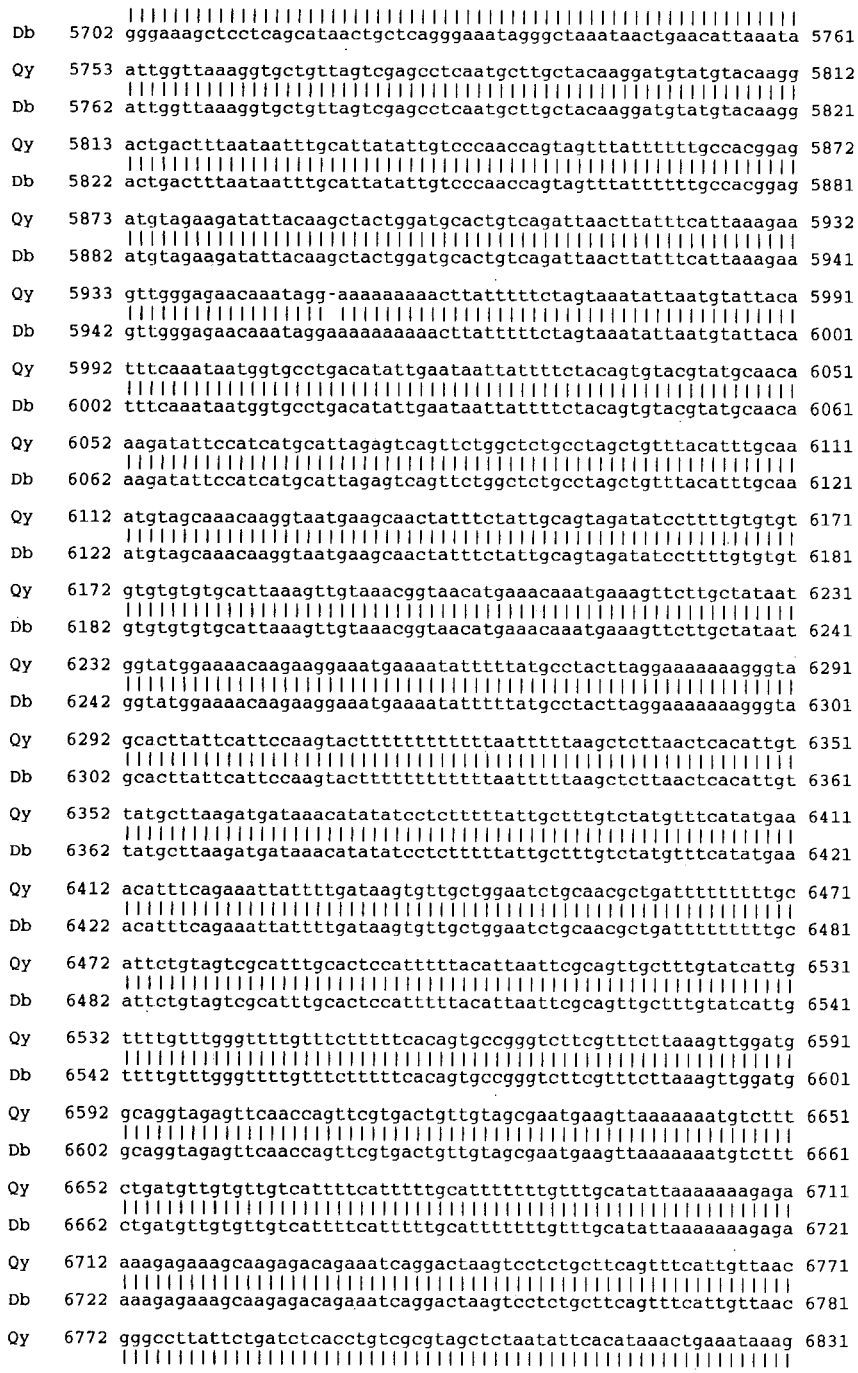
Qy   5453 tttcttcacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttcca 5512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5462 tttcttcacaaggtaacctgctgtatttatttatttttcttttggttaaatataatttcca 5521

Qy   5513 aactttgtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgt 5572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5522 aactttgtggtcaggcagcgtctaaggttacgttaccacagactgacagttggtatatgt 5581

Qy   5573 accagccaatcccttcattaaatgtatacagatttagttaagtagcattaaataggattc 5632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5582 accagccaatcccttcattaaatgtatacagatttagttaagtagcattaaataggattc 5641

Qy   5633 ttagaagtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaa 5692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5642 ttagaagtatgtcctcatagaacttttaatacttaaggctttgtaaaaactatccatgaa 5701

Qy   5693 gggaaagctcctcagcataactgctcagggaaatagggctaaataactgaacattaaata 5752
```


Db 5702 gggaaagctcctcagcataactgctcagggaaatagggctaaataactgaacattaaata 5761
Qy 5753 attggttaaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaagg 5812
Db 5762 attggttaaaggtgctgttagtcgagcctcaatgcttgctacaaggatgtatgtacaagg 5821
Qy 5813 actgactttaataatttgcattatattgtcccaaccagtagtttatttttgccacggag 5872
Db 5822 actgactttaataatttgcattatattgtcccaaccagtagtttatttttgccacggag 5881
Qy 5873 atgtagaagatattacaagctactggatgcactgtcagattaacttatttcattaaagaa 5932
Db 5882 atgtagaagatattacaagctactggatgcactgtcagattaacttatttcattaaagaa 5941
Qy 5933 gttgggagaacaaatagg-aaaaaaaaaacttatttttctagtaaatattaatgtattaca 5991
Db 5942 gttgggagaacaaatagggaaaaaaaaaacttatttttctagtaaatattaatgtattaca 6001
Qy 5992 tttcaaataatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaaca 6051
Db 6002 tttcaaataatggtgcctgacatattgaataattatttttctacagtgtacgtatgcaaca 6061
Qy 6052 aagatattccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaa 6111
Db 6062 aagatattccatcatgcattagagtcagttctggctctgcctagctgtttacatttgcaa 6121
Qy 6112 atgtagcaaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgt 6171
Db 6122 atgtagcaaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgt 6181
Qy 6172 gtgtgtgtgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataat 6231
Db 6182 gtgtgtgtgcattaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataat 6241
Qy 6232 ggtatggaaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggta 6291
Db 6242 ggtatggaaaacaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggta 6301
Qy 6292 gcacttattcattccaagtacttttttttttttaatttttaagctcttaactcacattgt 6351
Db 6302 gcacttattcattccaagtacttttttttttttaatttttaagctcttaactcacattgt 6361
Qy 6352 tatgcttaagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaa 6411
Db 6362 tatgcttaagatgataaacatatatcctctttttattgctttgtctatgtttcatatgaa 6421
Qy 6412 acatttcagaaattatttttgataagtgttgctggaatctgcaacgctgatttttttttgc 6471
Db 6422 acatttcagaaattatttttgataagtgttgctggaatctgcaacgctgatttttttttgc 6481
Qy 6472 attctgtagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattg 6531
Db 6482 attctgtagtcgcatttgcactccatttttacattaattcgcagttgctttgtatcattg 6541
Qy 6532 ttttgtttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatg 6591
Db 6542 ttttgtttgggttttgtttctttttcacagtgccgggtcttcgtttcttaaagttggatg 6601
Qy 6592 gcaggtagagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtcttt 6651
Db 6602 gcaggtagagttcaaccagttcgtgactgttgtagcgaatgaagttaaaaaaatgtcttt 6661
Qy 6652 ctgatgttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagaga 6711
Db 6662 ctgatgttgtgttgtcattttcatttttgcatttttttgtttgcatattaaaaaaagaga 6721
Qy 6712 aaagagaaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaac 6771
Db 6722 aaagagaaagcaagagacagaaatcaggactaagtcctctgcttcagtttcattgttaac 6781
Qy 6772 gggccttattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaag 6831
Db 6782 gggccttattctgatctcacctgtcgcgtagctctaatattcacataaactgaaataaag 6841
Qy 6832 aagtggaatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaat 6891
Db 6842 aagtggaatgaggagctttgacattcaaattatgtgatgtaatttatcttccttaggaat 6901
Qy 6892 tttgatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaa 6951
Db 6902 tttgatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaa 6961
Qy 6952 aaaagagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacag 7011
Db 6962 aaaagagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacag 7021
Qy 7012 aatgcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaac 7071
Db 7022 aatgcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaac 7081
Qy 7072 accagaagtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactc 7131
Db 7082 accagaagtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactc 7141
Qy 7132 attaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgc 7191
Db 7142 attaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgc 7201
Qy 7192 ctttttatttttcattcaatttttttttttttctcagacagacatagtagtatcaactagca 7251
Db 7202 ctttttatttttcattcaatttttttttttttctcagacagacatagtagtatcaactagca 7261
Qy 7252 ttggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaat 7311
Db 7262 ttggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaat 7321
Qy 7312 atttttggatagcgttgacacgatagatcttattccatacttctttattattgataattt 7371
Db 7322 atttttggatagcgttgacacgatagatcttattccatacttctttattattgataattt 7381
Qy 7372 tattttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttt 7431
Db 7382 tattttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttt 7441
Qy 7432 ttgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagcct 7491
Db 7442 ttgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagcct 7501
Qy 7492 ttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatt 7551
Db 7502 ttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatt 7561
Qy 7552 attttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgag 7611
Db 7562 attttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgag 7621
Qy 7612 gaatttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaat 7671
Db 7622 gaatttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaat 7681
Qy 7672 gatgtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtc 7731
Db 7682 gatgtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtc 7741
Qy 7732 tatattctaaattgagctttggatcaaacttttaggccaggaccagctcatgcgttctcat 7791
Db 7742 tatattctaaattgagctttggatcaaacttttaggccaggaccagctcatgcgttctcat 7801
Qy 7792 tcttcctttttctcactctttctctcatcactcacctctgtattcattctgttgtttggga 7851
Db 7802 tcttcctttttctcactctttctctcatcactcacctctgtattcattctgttgtttggga 7861
Qy 7852 tagaaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactaca 7911
Db 7862 tagaaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactaca 7921

```
Qy   7912 tgactccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg 7971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7922 tgactccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaagg 7981

Qy   7972 -aaagagtatgcccaatttctctctacatgacatattgagatttttttttaatcaacttttta 8030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7982 aaaagagtatgcccaatttctctctacatgacatattgagatttttttttaatcaacttttta 8041

Qy   8031 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaca 8090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8042 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaca 8101

Qy   8091 agcatggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacag 8150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8102 agcatggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacag 8161

Qy   8151 ctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatct 8210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8162 ctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatct 8221

Qy   8211 attatgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgc 8270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8222 attatgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgc 8281

Qy   8271 ctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataatt 8330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8282 ctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataatt 8341

Qy   8331 attagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtga 8390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8342 attagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtga 8401

Qy   8391 actgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatc 8450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8402 actgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatc 8461

Qy   8451 cagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcct 8510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8462 cagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcct 8521

Qy   8511 tttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 8570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8522 tttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 8581

Qy   8571 aaaaaaaaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagctt 8630
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   8582 aaaaaaaaaaaaaaaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagctt 8641

Qy   8631 gctgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaac 8690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8642 gctgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaac 8701

Qy   8691 taaatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaa 8750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8702 taaatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaa 8761

Qy   8751 atgagaagaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaa 8810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8762 atgagaagaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaa 8821

Qy   8811 gaaagacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttca 8870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8822 gaaagacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttca 8881

Qy   8871 gtactgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaag 8930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8882 gtactgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaag 8941

Qy   8931 catcccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaa 8990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8942 catcccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaa 9001

Qy   8991 gtatgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaa 9050
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9002 gtatgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaa 9061

Qy   9051 tgtacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctac 9110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9062 tgtacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctac 9121

Qy   9111 accggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaa 9170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9122 accggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatcagagaaa 9181

Qy   9171 attgtccttctagagcatagagtcccaaaactcaattctggttttcccctgttttttttt 9230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9182 attgtccttctagagcatagagtcccaaaactcaattctggttttcccctgttttttttt 9241

Qy   9231 ttttttttttcccaacatatgaactgcagcatatcacttttttcttttgtgcctcaggt 9290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9242 ttttttttttcccaacatatgaactgcagcatatcacttttttcttttgtgcctcaggt 9301

Qy   9291 tcctcacctgtaaaattgaaaaatatatgtattaataatattattaataataataatggt 9350
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9302 tcctcagctgtaaaattgaaaaatatatgtattaataatattattaataataataatggt 9361

Qy   9351 aatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtg 9410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9362 aatgtagtacttgtttgtaaagcactttgagatccttggttgaaaggcaccataggagtg 9421

Qy   9411 ccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaa 9470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9422 ccaagtattattatgtggccaaggggttatttaaactgtcagttcccaaaggccaggaa 9481

Qy   9471 aggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgt 9530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9482 aggttggggtcatttttcttaaagacgagctgtaaatatcaactaggcagccaatagtgt 9541

Qy   9531 tgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctac 9590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9542 tgactatgaagatgcaaaactattactaggctgataaaatcatagtttcttaatggctac 9601

Qy   9591 caataaggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaac 9650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9602 caataaggcaaatatcacaataataaacgccaaattccttagggcggactatttgacaac 9661

Qy   9651 cacatggaaaacttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaat 9710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9662 cacatggaaaacttgggggaggcatgaggggggaacatctcaaaatgccaatgtaaaat 9721

Qy   9711 ttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttg 9770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9722 ttaacttacagcaatattcaccagcagaaaatgtctttcatatggaatgatttcatgttg 9781

Qy   9771 ctaagaaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatag 9830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9782 ctaagaaaaagaattcaatttgtagtcctgatttgaatactagaatgttggctataatag 9841

Qy   9831 ttctgttcttacaacacatgaaatttttcgttttatttattttgttttcatagtgcat 9890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9842 ttctgttcttacaacacatgaaatttttcgttttatttattttgttttcatagtgcat 9901

Qy   9891 gttcatttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggca 9950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9902 gttcatttctactcacaaacatgttcttggtgtatttcttatgcaaacaatcttcaggca 9961

Qy   9951 gcaaagatgtctgttacatctaaacttgaataataaagtttt-ccaccagttacaca 10006
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   9962 gcaaagatgtctgttacatctaaacttgaataataaagttttaccaccagttacaca 10018
```

RESULT 3
C79479
ID C79479 standard; cDNA; 1009 BP.
XX

```
AC   C79479;
XX
DT   07-FEB-2001  (first entry)
XX
DE   cDNA sequence for human breast tumour clone #19606.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy;
KW   breast cancer; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09688.
XX
PR   09-APR-1999;   99US-0288950.
PR   02-JUL-1999;   99US-0346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer -
XX
PS   Claim 4; Page 95; 95pp; English.
XX
CC   The present sequence was isolated from a breast tumour cDNA library. It
CC   is provided in a specification relating to compounds for immunotherapy
CC   and diagnosis of breast cancer. Breast tumour antigens and the
CC   polynucleotides that encode them may be used in the production of a
CC   pharmaceutical composition to be used in the treatment of breast cancer.
CC   Proliferated T cells and incubated antigen presenting cells are also
CC   required. The polypeptides and polynucleotides may also be used to
CC   produce a vaccine.
XX
SQ   Sequence 1009 BP; 316 A; 163 C; 193 G; 336 T; 1 other;


  Query Match             9.9%;  Score 995;  DB 21;  Length 1009;
  Best Local Similarity   99.7%;  Pred. No. 2.6e-203;
  Matches 1006;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy   7578 cgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatttt 7637
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatttt 60

Qy   7638 ctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgcca 7697
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgcca 120

Qy   7698 tttaaaaactgaaaatattttaaaattatttgtctatattctaaattgagctttggatca 7757
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tttaaaaactgaaaatattttaaaattatttgtctatattctaaattgagctttggatca 180

Qy   7758 aactttaggccaggaccagctcatgcgttctcattcttccttttctcactctttctctca 7817
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aactttaggccaggaccagctcatgcgttctcattcttccttttctcactctttctctca 240

Qy   7818 tcactcacctctgtattcattctgttgtttgggatagaaaaatcataaagagccaaccca 7877
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcactcacctctgtattcattctgttgtttgggatagaaaaatcataaagagccaaccca 300

Qy   7878 tctcagaacgttgtggattgagagagacactacatgactccaagtatatgagaaaaggac 7937
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 tctcagaacgttgtggattgagagagacactacatgactccaagtatatgagaaaaggac 360

Qy   7938 agagctctaattgataactctgtagttcaaaagg-aaagagtatgcccaattctctctac 7996
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    361 agagctctaattgataactctgtagttcaaaaggaaaagagtatgcccaattctctctac 420

Qy   7997 atgacatattgagatttttttaatcaacttttaagatagtgatgttctgttctaaactg 8056
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 atgacatattgagatttttttaatcaacttttaagatagtgatgttctgttctaaactg 480

Qy   8057 ttctgttttagtgaaggtagatttttataaaaacaagcatggggattcttttctaaggtaa 8116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttctgttttagtgaaggtagatttttataaaacaagcatggggattcttttctaaggtaa 540

Qy   8117 tattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcac 8176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db    541 tattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcmc 600

Qy   8177 attatgtttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctc 8236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 attatgtttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctc 660

Qy   8237 caaaaatattaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataa 8296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 caaaaatattaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataa 720

Qy   8297 actgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaa 8356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 actgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaa 780

Qy   8357 aacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaa 8416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 aacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaa 840

Qy   8417 ttaatgcatttctagaggggggaatatctgccatccagtggtggaaatgtggagtaaagaa 8476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttaatgcatttctagaggggggaatatctgccatccagtggtggaaatgtggagtaaagaa 900

Qy   8477 gctggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaa 8536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 gctggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaa 960

Qy   8537 ctttagctactgtctttggtttgagagccatggcaaaaaaaaaaaaaga 8585
            |||||||||||||||||||||||||||||||||||||||||||||| |
Db    961 ctttagctactgtctttggtttgagagccatggcaaaaaaaaaaaaaa 1009


RESULT  4
F44852
ID   F44852 standard; cDNA; 996 BP.
XX
AC   F44852;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human breast cancer related protein coding sequence SEQ ID NO: 8.
XX
KW   Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078960-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17536.
XX
PR   23-JUN-1999;   99US-0140903.
PR   12-OCT-1999;   99US-0158980.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Mitcham JL;
XX
DR   WPI; 2001-041426/05.
XX
```

```
PT   New polynucleotides encoding breast tumor specific proteins, useful for
PT   prevention, treatment and diagnosis of breast cancer -
XX
PS   Claim 5; Page 120; 165pp; English.
XX
CC   The present invention provides the coding sequences for a number of
CC   breast cancer related proteins. These can be used in vaccinations
CC   against, diagnosis of and treatment of cancer, particularly breast
CC   cancer.
XX
SQ   Sequence 996 BP; 299 A; 148 C; 188 G; 361 T; 0 other;


  Query Match             9.3%;  Score 931.4;  DB 22;  Length 996;
  Best Local Similarity   99.6%;  Pred. No. 1e-189;
  Matches  965;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy   5891 gctactggatgcactgtcagattaacttatttcattaaagaagttgggagaacaaatagg 5950
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     29 gctactggatgcactgtcagattaacttatttcattaaagaagttgggagaacaaatagg 88

Qy   5951 -aaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaataatggtgcct 6009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 aaaaaaaaacttatttttctagtaaatattaatgtattacatttcaaataatggtgcct 148

Qy   6010 gacatattgaataattattttctacagtgtacgtatgcaacaaagatattccatcatgca 6069
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 gacatattgaataattattttctacagtgtacgtatgcaacaaagatattccatcatgca 208

Qy   6070 ttagagtcagttctggctctgcctagctgtttacatttgcaaatgtagcaaacaaggtaa 6129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 ttagagtcagttctggctctgcctagctgtttacatttgcaaatgtagcaaacaaggtaa 268

Qy   6130 tgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtgtgcattaaagt 6189
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 tgaagcaactatttctattgcagtagatatccttttgtgtgtgtgtgtgtgcattaaagt 328

Qy   6190 tgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatggaaaacaagaag 6249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 tgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatggaaaacaagaag 388

Qy   6250 gaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcacttattcattccaag 6309
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 gaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcacttattcattccaag 448

Qy   6310 tacttttttttttttaatttttaagctcttaactcacattgttatgcttaagatgataaa 6369
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 tac-ttttttttttaatttttaagctcttaactcacattgttatgcttaagatgataaa 507

Qy   6370 catatatcctcttttattgctttgtctatgtttcatatgaaacatttcagaaattattt 6429
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    508 catatatcctcttttattgctttgtctatgtttcatatgaaacatttcagaaattattt 567

Qy   6430 tgataagtgttgctggaatctgcaacgctgat-ttttttttgcattctgtagtcgcattt 6488
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    568 tgataagtgttgctggaatctgcaacgctgatgttttttttgcattctgtagtcgcattt 627

Qy   6489 gcactccatttttacattaattcgcagttgctttgtatcattgttttgtttgggttttgt 6548
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    628 gcactccatttttacattaattcgcagttgctttgtatcattgttttgtttgggttttgt 687

Qy   6549 ttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggtagagttcaacc 6608
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688 ttctttttcacagtgccgggtcttcgtttcttaaagttggatggcaggtagagttcaacc 747

Qy   6609 agttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatgttgtgttgtca 6668
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    748 agttcgtgactgttgtagcgaatgaagttaaaaaaatgtctttctgatgttgtgttgtca 807

Qy   6669 ttttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagagaaagcaagaga 6728
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    808 ttttcatttttgcatttttttgtttgcatattaaaaaaagagaaaagagaaagcaagaga 867

Qy   6729 cagaaatcaggactaagtcctctgcttcagtttcattgttaacgggccttattctgatct 6788
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    868 cagaaatcaggactaagtcctctgcttcagtttcattgttaacgggccttattctgatct 927

Qy   6789 cacctgtcgcgtagctctaatattcacataaactgaaataaagaagtggaatgaggagct 6848
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    928 cacctgtcgcgtagctctaatattcacataaactgaaataaagaagtggaatgaggagct 987

Qy   6849 ttgacattc 6857
            ||||||| ||
Db    988 ttgacagtc 996


RESULT   5
F44875
ID   F44875 standard; cDNA; 674 BP.
XX
AC   F44875;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human breast cancer related protein coding sequence SEQ ID NO: 31.
XX
KW   Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078960-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17536.
XX
PR   23-JUN-1999;   99US-0140903.
PR   12-OCT-1999;   99US-0158980.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Mitcham JL;
XX
DR   WPI; 2001-041426/05.
XX
PT   New polynucleotides encoding breast tumor specific proteins, useful for
PT   prevention, treatment and diagnosis of breast cancer -
XX
PS   Claim 5; Page 128; 165pp; English.
XX
CC   The present invention provides the coding sequences for a number of
CC   breast cancer related proteins. These can be used in vaccinations
CC   against, diagnosis of and treatment of cancer, particularly breast
CC   cancer.
XX
SQ   Sequence 674 BP; 222 A; 110 C; 128 G; 213 T; 1 other;


  Query Match             6.7%;  Score 668.8;  DB 22;  Length 674;
  Best Local Similarity   99.4%;  Pred. No. 1.3e-133;
  Matches  670;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   8958 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 9017
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 60

Qy   9018 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 9077
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 120

Qy   9078 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 9137
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 180

Qy   9138 cgtggaaataatgtttagctttttatcagagaaaattgtccttctagagcatagagtccca 9197
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181 cgtggaaataatgtttagcttttatcagagaaaattgtccttctagagcatagagtccca 240

Qy   9198 aaactcaattctggttttcccctgttttttttttttttttttttttcccaacatatgaactg 9257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aaactcaattctggttttcccctgttttttttttttttttttttttcccaacatatgaactg 300

Qy   9258 cagcatatcacttttttcttttttgtgcctcaggttcctcacctgtaaaattgaaaaatata 9317
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    301 cagcatatcacttttttcttttttgtgcctcaggttcctcagctgtaaaattgaaaaatata 360

Qy   9318 tgtattaataatattattaataataataatggtaatgtagtacttgtttgtaaagcactt 9377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tgtattaataatattattaataataataatggtaatgtagtacttgtttgtaaagcactt 420

Qy   9378 tgagatccttggttgaaaggcaccataggagtgccaagtattattatgtggccaaggggg 9437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgagatccttggttgaaaggcaccataggagtgccaagtattattatgtggccaaggggg 480

Qy   9438 ttatttaaactgtcagttcccaaaggccaggaaaggttggggtcatttttcttaaagacg 9497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttatttaaactgtcagttcccaaaggccaggaaaggttggggtcatttttcttaaagacg 540

Qy   9498 agctgtaaatatcaactaggcagccaatagtgttgactatgaagatgcaaaactattact 9557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 agctgtaaatatcaactaggcagccaatagtgttgactatgaagatgcaaaactattact 600

Qy   9558 aggctgataaaatcatagtttcttaatggctaccaataaggcaaatatcacaataataaa 9617
          |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 aggmtgataaaatcatagtttcttaatggctaccaataaggcaaatatcacaataataaa 660

Qy   9618 cgccaaattcctta 9631
          | |||||||||| ||
Db    661 caccaaattccgta 674
```

```
RESULT   6
C16523
ID   C16523 standard; cDNA; 357 BP.
XX
AC   C16523;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 20598.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 20598; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 357 BP; 110 A; 91 C; 73 G; 83 T; 0 other;
```

```
  Query Match             3.3%;  Score 333;  DB 21;  Length 357;
  Best Local Similarity   99.4%;  Pred. No. 6.2e-62;
  Matches  355;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy   1681 ccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctctccccct 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ccgaattagaacaacattttcttcagactcacccaaacaaaataaaagcttctctccccct 60

Qy   1741 cctctgaggttgcaaaaccttcagagaaaaactctaacaagtccatccctgcacttcaat 1800
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 cctctgaggttgc-aaaccttcagagaaaaactctaacaagtccatccctgcacttcaat 119

Qy   1801 ccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcaggagatg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 ccagtgattctggagacttgggaaaatggcaggacaagataacagtcaaagcaggagatg 179

Qy   1861 acactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaaa-ggta 1919
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    180 acactcctgttgggtactcagtgcccataaagcccctcgattcctctagacaaaatggta 239

Qy   1920 cagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcatctagct 1979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 cagaggccaccagttactactggtgtaaattttgtagtttcagctgtgagtcatctagct 299

Qy   1980 cacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcggcc 2036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 cacttaaactgctagaacattatggcaagcagcacggagcagtgcagtcaggcggcc 356
```

```
RESULT   7
V27248
ID   V27248 standard; DNA; 333 BP.
XX
AC   V27248;
XX
DT   28-SEP-1998  (first entry)
XX
DE   HeLa gene sequence E7_10_9.
XX
KW   Gene identification; HeLa; Fas antibody; cancer; antisense;
KW   therapy; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO9821366-A1.
XX
PD   22-MAY-1998.
XX
PF   12-NOV-1997;   97WO-US20989.
XX
PR   13-NOV-1996;   96US-0030549.
XX
PA   (QBIQ-) QBI ENTERPRISES LTD.
XX
PI   Deiss LP,  Efimova E,  Einat P,  Vasquez-Iaslop NC;
PI   Yehiely F;
XX
DR   WPI; 1998-297966/26.
```

```
XX
PT   Identifying genes essential for maintenance of specific cell
PT   phenotypes - using gene inactivation, positive selection and
PT   subtraction analysis, useful to develop pharmaceuticals to undesired
PT   phenotypes e.g. cancer
XX
PS   Example 2a; Page 51; 66pp; English.
XX
CC   E7_10_9 is one of 22 gene sequences (see V27227-48) identified
CC   in HeLa cells as being involved in Fas antibody sensitivity.  A
CC   novel method of the invention was applied to HeLa cells treated
CC   with anti-Fas antibody in order to identify genes that when
CC   knocked-out cause sensitisation of HeLa cells to the action of
CC   anti-Fas antibodies.  E7_10_9 shows no good match against known
CC   genes in a combined nonredundant database, but does show good
CC   matches against 2 ESTs (gb R54192 and gb H39863).  Sequence
CC   analysis indicates that the fragment inserted in the expression
CC   library was in the antisense orientation.  Thus, reduction of
CC   the gene product that is encoded by the full-length message
CC   representing this clone leads to supersensitisation of cells to
CC   treatment with anti-Fas antibodies.
XX
SQ   Sequence 333 BP; 100 A; 58 C; 83 G; 92 T; 0 other;


  Query Match             3.2%;  Score 315.4;  DB 19;  Length 333;
  Best Local Similarity   99.7%;  Pred. No. 3.5e-58;
  Matches  316;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    8780 ggactttggagggcattgttatttcccaaagaaagacggccaagggcagaggcatggat 8839
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       5 ggactttggagggcagtgttatttcccaaagaaagacggccaagggcagaggcatggat 64

Qy    8840 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 8899
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 124

Qy    8900 gttcctgatagtgctgcagttgcttagaaagcatcccagttaattgcagtaattagaaact 8959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 gttcctgatagtgctgcagttgcttagaaagcatcccagttaattgcagtaattagaaact 184

Qy    8960 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 9019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 244

Qy    9020 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 9079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 304

Qy    9080 agggagagctgtaagct 9096
           |||||||||||||||||
Db     305 agggagagctgtaagct 321


RESULT   8
C08664
ID   C08664 standard; cDNA; 187 BP.
XX
AC   C08664;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 12739.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12739; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 187 BP; 35 A; 46 C; 38 G; 68 T; 0 other;


  Query Match             1.9%;  Score 187;  DB 21;  Length 187;
  Best Local Similarity   100.0%;  Pred. No. 8e-31;
  Matches  187;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      22 gatattaatagtgttggtgtcttgaaactgacgtaatgcgcggagactgaggtcctgaca 81
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gatattaatagtgttggtgtcttgaaactgacgtaatgcgcggagactgaggtcctgaca 60

Qy      82 agcgataacatttctgataaagacccgatcttactgcaatctctagcgtcctcttttttg 141
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 agcgataacatttctgataaagacccgatcttactgcaatctctagcgtcctcttttttg 120

Qy     142 gtgctgctggtttctccagacctcgcgtcctctcgattgctctctcgccttcctatttct 201
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gtgctgctggtttctccagacctcgcgtcctctcgattgctctctcgccttcctatttct 180

Qy     202 ttttttt 208
           |||||||
Db     181 ttttttt 187


RESULT   9
C21536
ID   C21536 standard; cDNA; 173 BP.
XX
AC   C21536;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 25611.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
```

```
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 25611; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 173 BP; 58 A; 25 C; 30 G; 55 T; 5 other;


  Query Match             1.7%;  Score 169.4;  DB 21;  Length 173;
  Best Local Similarity   96.5%;  Pred. No. 4.5e-27;
  Matches  167;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    8067 gtgaaggtagatttttataaaacaagcatggggattcttttctaaggtaatattaatgag 8126
           ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       1 gtgaaggtagatttttatamaacaagcatggggattcttttctaaggtaatattaatgag 60

Qy    8127 aagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcacattatgttta 8186
           |||||||||||||||||||||||||||||:||||||||||:||||||||||||||||||
Db      61 aagggaaaaaagtatctttaacagctctttrttgaagcctgtrgtagcacattatgttta 120

Qy    8187 taattgcacatgtgcacataatctattatgatccaatgcaaatacagctccaa 8239
           |||||:|||||||||||||||||||||||||||||||||||||||||| |||:
Db     121 taattrcacatgtgcacataatctattatgatccaatgcaaatacagccccar 173


RESULT  10
C28768
ID   C28768 standard; cDNA; 76 BP.
XX
AC   C28768;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 32843.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 32843; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 76 BP; 25 A; 12 C; 17 G; 22 T; 0 other;


  Query Match             0.8%;  Score 76;  DB 21;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 3e-07;
  Matches   76;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    8508 ccttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccat 8567
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ccttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccat 60

Qy    8568 ggcaaaaaaaaaaaaa 8583
           ||||||||||||||||
Db      61 ggcaaaaaaaaaaaaa 76


RESULT  11
T72882/c
ID   T72882 standard; cDNA; 19124 BP.
XX
AC   T72882;
XX
DT   12-SEP-1997  (first entry)
XX
DE   Plasmodium var-7 gene.
XX
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
XX
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   exon            7317..15139
FT                   /*tag=  a
FT                   /number= 1
FT   intron          15140..16205
FT                   /*tag=  b
FT                   /number= 1
FT   exon            16206..17552
FT                   /*tag=  c
FT                   /number= 2
```

```
FT                   /note= "no stop codon given"
XX
PN   WO9640766-A2.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09508.
XX
PR   07-JUN-1995;   95US-0487826.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Chitnis C,  Miller LH,  Peterson DS,  Sim KL,  Su X;
PI   Wellems TE;
XX
DR   WPI; 1997-052231/05.
DR   P-PSDB; W22475.
XX
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
XX
PS   Claim 4; Page 56-61; 96pp; English.
XX
CC   This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and
CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
XX
SQ   Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

  Query Match             0.7%;  Score 72.4;  DB 18;  Length 19124;
  Best Local Similarity   49.0%;  Pred. No. 1.6e-05;
  Matches  257;  Conservative    0;  Mismatches  256;  Indels   11;  Gaps    2;

Qy    7057 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattactttt 7116
           |||  |||  |||        || |||  || | ||    ||  |       | |    |||
Db   15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy    7117 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 7176
           |     |    | | | |       ||   || | | |   |   |||||  || ||| |
Db   15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy    7177 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 7236
            |||     | |||   ||||| ||  ||||   ||  ||||||||          |  |
Db   15800 ATTTTATGTATATATTTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy    7237 tagtatcaactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 7296
           |  |||  | |   |||  || |||  | |  || || ||    | ||||||  |   | |
Db   15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy    7297 acttttttagtaaaatatttttggatagcgttgacacgatagatcttattccatac---tt 7353
             |||||  |  ||| |||||   |  | |         |  || || ||    |    ||
Db   15688 TTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTTTATAATCATTT 15629

Qy    7354 ctttattattgataatttttattttcatttttgctttcattattatacatattttggtgg 7413
            ||| ||||  | |||||| |||| |||||||  |   | | || | ||| ||||  |

Db   15628 TTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569

Qy    7414 agaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttt 7473
           |  | |  ||    ||| ||  ||      |  | ||| | | ||   |     ||  |||
Db   15568 ATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy    7474 tttgacatattaaagcctttattccatctctcaagatatattataaaatttatttttttta 7533
           |||    | ||      |||| |  || | |    |||| | ||     || ||||  |
Db   15508 TTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy    7534 atttaagatttctgaattattttatcttaaattgtgattttaaa 7577
            |   | ||||  |   |   ||| | ||  | | | | |  | |
Db   15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405


RESULT  12
Z98287/c
ID   Z98287 standard; DNA; 19124 BP.
XX
AC   Z98287;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Plasmodium var-7 polypeptide encoding DNA.
XX
KW   DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW   DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
KW   protozoacide; var-7; ds.
XX
OS   Plasmodium sp.
XX
PN   US5993827-A.
XX
PD   30-NOV-1999.
XX
PF   07-JUN-1995;   95US-0487826.
XX
PR   10-SEP-1993;   93US-0119677.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Sim KL,  Chitnis C,  Peterson DS,  Su X,  Wellems TE,  Miller LH;
XX
DR   WPI; 2000-194198/17.
DR   P-PSDB; Y77904.
XX
PT   Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT   falciparum erythrocyte binding proteins useful for vaccinating against
PT   malaria -
XX
PS   Disclosure; Columns 91-108; 93pp; English.
XX
CC   The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC   (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC   identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC   Binding Protein (SABP), which are soluble proteins that appear in the
CC   culture supernatant after erythrocytes infected with malaria release
CC   merozoites. Immunochemical studies indicate that DABP and SABP are the
CC   respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC   and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC   used to vaccinate against malaria, especially caused by P. falciparum.
CC   Immunization with the polypeptide provides effective protection against
CC   malaria. The present sequence represents the var-7 polypeptide encoding
CC   DNA.
XX
SQ   Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

  Query Match             0.7%;  Score 72.4;  DB 21;  Length 19124;
  Best Local Similarity   49.0%;  Pred. No. 1.6e-05;
  Matches  257;  Conservative    0;  Mismatches  256;  Indels   11;  Gaps    2;

Qy    7057 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattactttt 7116
```

```
Db  15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy   7117 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 7176
Db  15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy   7177 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 7236
Db  15800 ATTTTATGTATATATTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy   7237 tagtatcaactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 7296
Db  15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy   7297 acttttttagtaaaatatttttggatagcgttgacacgatagatcttattccatac---tt 7353
Db  15688 TTTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTTATAATCATTT 15629

Qy   7354 ctttattattgataattttattttcatttttttgctttcattattatacatattttggtgg 7413
Db  15628 TTTTTTTATATAAAATTTTTTTTTTAATTTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569

Qy   7414 agaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttt 7473
Db  15568 ATCAAAATTTATATTTTATTATAATTTTTTATTATTTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy   7474 tttgacatattaaagccttttattccatctctcaagatatattataaaatttatttttttta 7533
Db  15508 TTTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy   7534 atttaagatttctgaattattttatcttaaattgtgattttaaa 7577
Db  15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405


RESULT 13
X33181/c
ID   X33181 standard; DNA; 6644 BP.
XX
AC   X33181;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-ires-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Cowpox virus.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP04010.
XX
PR   08-SEP-1997;   97JP-0259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell
XX
PS   Example 1; Page 38-41; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC   is used in an example from the present invention.
XX
SQ   Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;


  Query Match              0.7%;  Score 70.6;  DB 20;  Length 6644;
  Best Local Similarity   47.1%;  Pred. No. 2.5e-05;
  Matches  217;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy   7169 ttctttatttttggcatgtatgccttttttatttcattcaatttttttttttttctcagac 7228
Db   4180 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4121

Qy   7229 agacatagtagtatcaactagcattggaaaatacatatcactattcttggaatatttatg 7288
Db   4120 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4061

Qy   7289 gtcagtctacttttttagtaaaatatttttggatagcgttgacacgatagatcttattcca 7348
Db   4060 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4001

Qy   7349 tacttctttattattgataatttttattttcatttttttgctttcattattatacatatttt 7408
Db   4000 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3941

Qy   7409 ggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacact 7468
Db   3940 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3881

Qy   7469 cctttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 7528
Db   3880 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3821

Qy   7529 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 7588
Db   3820 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3761

Qy   7589 ggtacggaacttttttttaatgaggaatttcatgatgattta 7629
Db   3760 TTTTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 3720


RESULT 14
X33182/c
ID   X33182 standard; DNA; 7372 BP.
XX
AC   X33182;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-Bcl-xl-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
```

KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04010.
XX
PR 08-SEP-1997; 97JP-0259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell
XX
PS Example 2; Page 41-45; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and
CC is used in an example from the present invention.
XX
SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 other;

Query Match 0.7%; Score 70.6; DB 20; Length 7372;
Best Local Similarity 47.1%; Pred. No. 2.6e-05;
Matches 217; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

Qy 7169 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 7228
Db 4908 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4849

Qy 7229 agacatagtagtatcaactagcattggaaaatacatatcactattcttggaatatttatg 7288
Db 4848 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4789

Qy 7289 gtcagtctacttttttagtaaaatatttttggatagcgttgacacgatagatcttattcca 7348
Db 4788 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4729

Qy 7349 tacttctttattattgataatttttattttcatttttttgctttcattattatacatatttt 7408
Db 4728 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4669

Qy 7409 ggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacact 7468
Db 4668 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4609

Qy 7469 cctttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 7528
Db 4608 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4549

Qy 7529 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 7588
Db 4548 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4489

Qy 7589 ggtacggaacttttttttaatgaggaatttcatgatgattta 7629
Db 4488 TTTTTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 4448

RESULT 15
X33180/c
ID X33180 standard; DNA; 7797 BP.
XX
AC X33180;
XX
DT 25-JUN-1999 (first entry)
XX
DE Cowpox virus bsr full length gene sequence.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Cowpox virus.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04010.
XX
PR 08-SEP-1997; 97JP-0259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell
XX
PS Example 1; Page 34-38; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the cowpox virus bsr gene which
CC is used in an example from the present invention.
XX
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;

Query Match 0.7%; Score 70.6; DB 20; Length 7797;

Best Local Similarity 47.1%; Pred. No. 2.7e-05;
Matches 217; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

```
Qy   7169 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 7228
          || ||| |||||    | | |   ||||| |||| ||   ||||||||||| |
Db   5333 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5274

Qy   7229 agacatagtagtatcaactagcattggaaaatacatatcactattcttggaatatttatg 7288
            |  |  | |    |    ||       |   | |   | || ||    | ||| |
Db   5273 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5214

Qy   7289 gtcagtctacttttttagtaaaatatttttggatagcgttgacacgatagatcttattcca 7348
           |   | |  |||||  |    | |||||   |    ||       |   | || ||
Db   5213 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5154

Qy   7349 tacttctttattattgataattttattttcatttttgctttcattattatacatatttt 7408
          |  || ||| || ||  |  |||| |||| ||||||  |||  || || |   | ||||
Db   5153 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5094

Qy   7409 ggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacact 7468
            |          ||    ||||||              ||| || |     |       |
Db   5093 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5034

Qy   7469 cctttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 7528
           ||||||    | ||      ||| ||   | | |     | | || |    ||| |||
Db   5033 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4974

Qy   7529 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 7588
          ||||  |||    ||| |   || |||| | ||   || |  ||||        | || |
Db   4973 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4914

Qy   7589 ggtacggaacttttttttaatgaggaatttcatgatgattta 7629
            |       |||||||  |  | || | || || | ||||
Db   4913 TTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 4873
```

Search completed: April 30, 2001, 19:18:08
Job time: 37502 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:34:00 ; Search time 330.07 Seconds
(without alignments)
5293.042 Million cell updates/sec

Title: US-09-721-183-20
Perfect score: 10006
Sequence: 1 ttcctccgcgaaggctcctt.....agttttccaccagttacaca 10006

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	315.4	3.2	333	3	US-09-284-782-36	Sequence 36, Appl
c 2	72.4	0.7	19124	2	US-08-487-826B-13	Sequence 13, Appl
3	69.2	0.7	7218	1	US-08-232-463-14	Sequence 14, Appl
4	65	0.6	1456	4	US-09-037-135-1	Sequence 1, Appli
c 5	58.2	0.6	6243	2	US-09-056-075-1	Sequence 1, Appli
6	57.4	0.6	1004	1	US-08-465-590-8	Sequence 8, Appli
7	57.4	0.6	1004	4	US-08-283-300A-7	Sequence 7, Appli
8	57.4	0.6	1004	5	PCT-US95-09345-7	Sequence 7, Appli
9	57.4	0.6	1386	1	US-08-465-590-3	Sequence 3, Appli
10	57.4	0.6	1386	4	US-08-283-300A-2	Sequence 2, Appli
11	57.4	0.6	1386	5	PCT-US95-09345-2	Sequence 2, Appli
12	57.4	0.6	1611	5	PCT-US93-08743-3	Sequence 3, Appli
13	57.4	0.6	5852	1	US-07-867-106-2	Sequence 2, Appli
14	56.8	0.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
c 15	54	0.5	7218	1	US-08-232-463-14	Sequence 14, Appl
16	53.8	0.5	2897	2	US-08-927-394-1	Sequence 1, Appli
17	53.2	0.5	731	1	US-08-451-405A-2	Sequence 2, Appli
c 18	51.2	0.5	2223	1	US-08-257-073-4	Sequence 4, Appli
19	50.6	0.5	1128	1	US-08-465-590-7	Sequence 7, Appli
20	50.6	0.5	1128	4	US-08-283-300A-6	Sequence 6, Appli
21	50.6	0.5	1128	5	PCT-US95-09345-6	Sequence 6, Appli
22	50.6	0.5	1170	1	US-08-465-590-6	Sequence 6, Appli
23	50.6	0.5	1170	4	US-08-283-300A-5	Sequence 5, Appli
24	50.6	0.5	1170	5	PCT-US95-09345-5	Sequence 5, Appli
25	50.6	0.5	1296	1	US-08-465-590-4	Sequence 4, Appli
26	50.6	0.5	1296	4	US-08-283-300A-3	Sequence 3, Appli
27	50.6	0.5	1296	5	PCT-US95-09345-3	Sequence 3, Appli
28	50.6	0.5	1788	1	US-08-465-590-2	Sequence 2, Appli
29	50.6	0.5	1788	4	US-08-283-300A-1	Sequence 1, Appli
30	50.6	0.5	1788	5	PCT-US93-08743-2	Sequence 2, Appli
31	50.6	0.5	1788	5	PCT-US95-09345-1	Sequence 1, Appli
32	50.6	0.5	2049	1	US-08-465-590-5	Sequence 5, Appli
33	50.6	0.5	2049	4	US-08-283-300A-4	Sequence 4, Appli
34	50.6	0.5	2049	5	PCT-US95-09345-4	Sequence 4, Appli
c 35	50.6	0.5	6152	4	US-08-973-462-1	Sequence 1, Appli
c 36	49	0.5	4838	4	US-08-852-629-15	Sequence 15, Appl
c 37	48.4	0.5	1511	1	US-07-991-867B-8	Sequence 8, Appli
c 38	48.4	0.5	1511	1	US-08-107-755A-8	Sequence 8, Appli
c 39	48.4	0.5	1511	2	US-08-544-332-8	Sequence 8, Appli
c 40	47.8	0.5	5433	4	US-08-929-329-1	Sequence 1, Appli
c 41	47.6	0.5	5852	1	US-07-867-106-2	Sequence 2, Appli
c 42	47.6	0.5	10395	1	US-08-245-809-3	Sequence 3, Appli
c 43	47.6	0.5	10396	1	US-08-245-809-5	Sequence 5, Appli
c 44	47.6	0.5	10798	1	US-08-107-748-2	Sequence 2, Appli
c 45	47.6	0.5	10798	5	PCT-US92-01385-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
US-09-284-782-36
; Sequence 36, Application US/09284782
; Patent No. 6057111
;  GENERAL INFORMATION:
;    APPLICANT:  ENTERPRISES, LTD., QBI
;    APPLICANT:  Deiss, Louis P.
;    APPLICANT:  Yehiely, Fruma
;    APPLICANT:  Efimova, Elena
;    APPLICANT:  Vasquez-Iaslop, No. 6057111a C.
;    APPLICANT:  Einat, Paz
;    TITLE OF INVENTION:  GENE IDENTIFICATION METHOD
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 6057111thwestern Highway, Suite 410
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/284,782
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Montgomery, Ilene N.
;      REGISTRATION NUMBER:  38,972
;      REFERENCE/DOCKET NUMBER:  0168-00022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (248) 539-5050
;      TELEFAX:  (248) 539-5055
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-284-782-36


  Query Match             3.2%;  Score 315.4;  DB 3;  Length 333;
  Best Local Similarity   99.7%;  Pred. No. 6.4e-69;
  Matches  316;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy   8780 ggactttggagggcattgttatttttcccaaagaaagacggccaagggcagaggcatggat 8839
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      5 GGACTTTGGAGGGCAGTGTTATTTTCCCAAAGAAAGACGGCCAAGGGCAGAGGCATGGAT 64

Qy   8840 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 8899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 TCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACTGTTTCATAGACAGTGGGCTCACAT 124

Qy   8900 gttcctgatagtgctgcagttgcttagaaagcatcccagttaattgcagtaattagaact 8959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCCCAGTTAATTGCAGTAATTAGAACT 184

Qy   8960 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 9019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 TCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGTCACATGAAGAAAATGTGAAATTC 244

Qy   9020 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 9079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 AAGAGTAATCCACACGTGAGAAACTAGACAATGTACATTCATGTGTTCTCTTGAAAGGAA 304

Qy   9080 agggagagctgtaagct 9096
          |||||||||||||||||
Db    305 AGGGAGAGCTGTAAGCT 321
```

```
RESULT   2
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13
```

```
  Query Match             0.7%;  Score 72.4;  DB 2;  Length 19124;
  Best Local Similarity   49.0%;  Pred. No. 1.1e-07;
  Matches  257;  Conservative    0;  Mismatches  256;  Indels   11;  Gaps    2;

Qy   7057 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattacttttt 7116
          ||| ||| |||       .|| ||| || | ||    || |     | |    |||
Db  15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy   7117 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 7176
          |     |    | | | |       ||   || | | |   |   |||||  || ||| |
Db  15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy   7177 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 7236
           |||    | |||   ||||| ||  ||||   ||  ||||||||         | |
Db  15800 ATTTTATGTATATATTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy   7237 tagtatcaactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 7296
          |  |||  | |   |||  || |||  | |  || || ||    | ||||||  |   | |
Db  15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy   7297 acttttttagtaaaatatttttggatagcgttgacacgatagatcttattccatac---tt 7353
            |||||  |  ||| |||||   |  | |        |  || || ||    |    ||
Db  15688 TTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTTATAATCATTT 15629

Qy   7354 ctttattattgataatttttattttcatttttttgctttcattattatacatattttggtgg 7413
          ||| ||||  | ||||||| |||| |||||||  |   | | || | ||| |||| |
Db  15628 TTTTTTTATATAAAATTTTTTTTTTAATTTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569

Qy   7414 agaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttt 7473
          |  | |  ||   ||| ||  ||     |  | ||| | | ||   |    ||  |||
Db  15568 ATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy   7474 tttgacatattaaagcctttattccatctctcaagatatattataaaatttatttttttta 7533
          |||    | ||      |||| |  || | |    |||| | ||     || ||||  |
Db  15508 TTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy   7534 atttaagatttctgaattattttatcttaaattgtgattttaaa 7577
          |   | |||| |   |   ||| | ||  | | | | |  | |
Db  15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405
```

```
RESULT   3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14


  Query Match             0.7%;  Score 69.2;  DB 1;  Length 7218;
  Best Local Similarity   2.2%;  Pred. No. 4e-07;
  Matches    5;  Conservative  165;  Mismatches   58;  Indels    0;  Gaps    0;

Qy      83 gcgataacatttctgataaagacccgatcttactgcaatctctagcgtcctcttttttgg 142
           |||||  : :::::  :     :::  :::: :: :  :::::  : ::::::::::::
Db    1062 GCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1121

Qy     143 tgctgctggtttctccagacctcgcgtcctctcgattgctctctcgccttcctatttctt 202
           : :: ::  :::::::   :::: : :::::::  :: ::::::: ::::::: ::::::
Db    1122 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1181

Qy     203 ttttttttttttaaacaaaaaacaacacccccctcccctctcccacccggcaccgggcaca 262
           :::::::::::::   :       :  : :::::::::::::::: :::  : ::   : :
Db    1182 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1241

Qy     263 tccttgctctatttcctttctctttctctctctctctctctttttt 310
           ::::: :::: :::::::::::::::::::::::::::::::::::
Db    1242 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1289


RESULT   4
US-09-037-135-1
; Sequence 1, Application US/09037135
; Patent No. 6107034
; GENERAL INFORMATION:
;  APPLICANT: Weigel, Ronald
;  TITLE OF INVENTION: GATA-3 EXPRESSION IN HUMAN BREAST
;  TITLE OF INVENTION: CARCINOMA
;  FILE REFERENCE: SUN-76P
;  CURRENT APPLICATION NUMBER: US/09/037,135
;  CURRENT FILING DATE: 1998-03-09
;  EARLIER APPLICATION NUMBER: 09/037,135
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1456
;   TYPE: DNA
;   ORGANISM: H. sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (120)...(1451)
US-09-037-135-1


  Query Match             0.6%;  Score 65;  DB 4;  Length 1456;
  Best Local Similarity   57.7%;  Pred. No. 1.8e-06;
  Matches  116;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy    3257 ggagaaaacaagtccaaggatgaatcccagtccctgttacggaggcgtagaggctccggt 3316
           ||| | ||| | ||    | |  || ||    | |          ||  |  | ||
Db    1005 ggacagaaccggcccctcattaagcccaagcgaaggctgtctgcagccaggagagcaggg 1064

Qy    3317 gttttttgtgccaattgcctgaccacaaagacctctctctggcgaaagaatgcaaatggc 3376
              |  ||||| || || | |||||| |  ||| | |||||| | | |||||| |||||
Db    1065 acgtcctgtgcgaactgtcagaccaccacaaccacactctggaggaggaatgccaatggg 1124

Qy    3377 ggatatgtatgcaacgcgtgtggcctctaccagaagcttcactcgactcccaggccttta 3436
           |    ||| ||||| || ||||| |||||| | |||||||||   | |  ||| ||  |
Db    1125 gaccctgtctgcaatgcctgtgggctctactacaagcttcacaatattaacagacccctg 1184

Qy    3437 aacatcattaaacaaaacaac 3457
           |  || |  ||  ||  || |
Db    1185 actatgaagaaggaaggcatc 1205


RESULT   5
US-09-056-075-1/c
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.95238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3770..4013
;      OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;      OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1


  Query Match             0.6%;  Score 58.2;  DB 2;  Length 6243;
  Best Local Similarity   46.6%;  Pred. No. 0.0002;
  Matches  186;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy    6182 attaaagttgtaaacggtaacatgaaacaaatgaaagttcttgctataatggtatggaaa 6241
           || ||||   | ||  |  ||        |||  ||||   || || |||||  ||     |
Db    1598 ATAAAAGCCCTTAAAGAAAAAGGCTTTAAAAATAAAGAAATTACTCTAATGTTAGAAGTA 1539

Qy    6242 acaagaaggaaatgaaaatatttttatgcctacttaggaaaaaaagggtagcacttattc 6301
            ||| ||  |  || |  |  |  | | ||  |   ||||| |||      | |
Db    1538 CCAATAAAAACATTAGAACGTCATATTACGTATATGAAAAAAAATGGGCTTTTATAAAGG 1479
```

```
Qy   6302 attccaagtacttttttttttttaatttttaagctcttaactcacattgttatgcttaag 6361
Db   1478 CTCATTTTTTATATTCTTTTCTTCAAAGATTATATAATATAAAAAAATTTTTTTCAAACT 1419

Qy   6362 atgataaacatatatcctctttttattgctttgtctatgtttcatatgaaacatttcaga 6421
Db   1418 TTAAATAAAAAATATTTTTATATTTTTTTATTTTTTTATTTTTATATTTTTTTATTTTTT 1359

Qy   6422 aattattttgataagtgttgctggaatctgcaacgctgatttttttttgcattctgtagt 6481
Db   1358 TATTTTTATATTTTTTTATTTTTATATTTTTTTATTTTTATATTTTTTTATTTTTATATT 1299

Qy   6482 cgcatttgcactccatttttacattaattcgcagttgctttgtatcattgttttgtttgg 6541
Db   1298 TTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTATTTTTT 1239

Qy   6542 gttttgtttctttttcacagtgccgggtcttcgtttctt 6580
Db   1238 TATTTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTT 1200
```

```
RESULT   6
US-08-465-590-8
; Sequence 8, Application US/08465590
; Patent No. 5824770
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS: A T CELL PATHWAY REGULATORY GENE
;    NUMBER OF SEQUENCES:  164
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUCETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,590
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MPG-006C2DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1002
US-08-465-590-8
```

```
  Query Match              0.6%;  Score 57.4;  DB 1;  Length 1004;
  Best Local Similarity   60.6%;  Pred. No. 0.00011;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
Db    832 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 891

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
Db    892 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 951

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
Db    952 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 986
```

```
RESULT   7
US-08-283-300A-7
; Sequence 7, Application US/08283300A
; Patent No. 6172278
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS TRANSGENIC CELLS AND ANIMALS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/283,300A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MGP-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1004
US-08-283-300A-7
```

```
  Query Match              0.6%;  Score 57.4;  DB 4;  Length 1004;
  Best Local Similarity   60.6%;  Pred. No. 0.00011;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;
```

```
Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  |  |||| |||||| | || |||||  |  | || ||| |
Db    832 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 891

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
          ||||| ||   | |||  |||||  |||||| ||| |||    | |   ||  ||||
Db    892 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 951

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db    952 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 986
```

```
RESULT   8
PCT-US95-09345-7
; Sequence 7, Application PC/TUS9509345
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS TRANSGENIC CELLS AND ANIMALS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09345
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/283,300
;      FILING DATE:  29-JULY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MGP-027PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1004
PCT-US95-09345-7


  Query Match             0.6%;  Score 57.4;  DB 5;  Length 1004;
  Best Local Similarity   60.6%;  Pred. No. 0.00011;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  |  |||| |||||| | || |||||  |  | || ||| |
Db    832 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 891

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
          ||||| ||   | |||  |||||  |||||| ||| |||    | |   ||  ||||
Db    892 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 951

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db    952 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 986


RESULT   9
US-08-465-590-3
; Sequence 3, Application US/08465590
; Patent No. 5824770
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS: A T CELL PATHWAY REGULATORY GENE
;    NUMBER OF SEQUENCES:  164
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUCETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,590
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MPG-006C2DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1386
US-08-465-590-3


  Query Match             0.6%;  Score 57.4;  DB 1;  Length 1386;
  Best Local Similarity   60.6%;  Pred. No. 0.00013;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  |  |||| |||||| | || |||||  |  | || ||| |
Db   1213 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 1272

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
```

```
          ||||| ||   | |||  |||||  |||||| ||| |||    | |   ||  ||||
Db   1273 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 1332

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db   1333 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 1367


RESULT 10
US-08-283-300A-2
; Sequence 2, Application US/08283300A
; Patent No. 6172278
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS TRANSGENIC CELLS AND ANIMALS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/283,300A
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MGP-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1386
US-08-283-300A-2


  Query Match             0.6%;  Score 57.4;  DB 4;  Length 1386;
  Best Local Similarity   60.6%;  Pred. No. 0.00013;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  | |||| |||||| | || |||||  |  | || ||| |
Db   1213 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 1272

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
          |||||  ||   | |||  |||||  |||||| ||| |||    | |   ||  ||||
Db   1273 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 1332

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db   1333 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 1367


RESULT 11
PCT-US95-09345-2
; Sequence 2, Application PC/TUS9509345
;  GENERAL INFORMATION:
;    APPLICANT:  Georgopoulos, Katia A.
;    TITLE OF INVENTION:  IKAROS TRANSGENIC CELLS AND ANIMALS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, Suite 510
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09345
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/283,300
;      FILING DATE:  29-JULY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/238,212
;      FILING DATE:  02-MAY-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,438
;      FILING DATE:  14-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/946,233
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,695
;      REFERENCE/DOCKET NUMBER:  MGP-027PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1386
PCT-US95-09345-2


  Query Match             0.6%;  Score 57.4;  DB 5;  Length 1386;
  Best Local Similarity   60.6%;  Pred. No. 0.00013;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  | |||| |||||| | || |||||  |  | || ||| |
Db   1213 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 1272

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
          |||||  ||   | |||  |||||  |||||| ||| |||    | |   ||  ||||
Db   1273 TGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGG 1332

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db   1333 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 1367
```

```
RESULT 12
PCT-US93-08743-3
; Sequence 3, Application PC/TUS9308743
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  IKAROS: A T CELL PATHWAY REGULATORY GENE
;    NUMBER OF SEQUENCES:  152
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08743
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 946,233
;      FILING DATE:  14-SEP-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1611 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1611
PCT-US93-08743-3


  Query Match             0.6%;  Score 57.4;  DB 5;  Length 1611;
  Best Local Similarity   60.6%;  Pred. No. 0.00015;
  Matches   94;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy   4282 aaatgtgtgcactgtggcattgtctttctggatgaagtgatgtatgctttgcatatgagt 4341
          || || |  |||||  |  |  |||| |||||| | || |||||  |  | || ||| |
Db   1435 AAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGC 1494

Qy   4342 tgccatggtgacagtggacctttccagtgcagcatatgccagcatctttgcacggacaaa 4401
          ||||| ||   | |||  ||||| |||||| ||| |||    | |   ||  ||||
Db   1495 TGCCACGGCTTCCGTGATCCTTTTGAGTGCACCATGTGCGGCTACCACAGCCAGGACCGG 1554

Qy   4402 tatgacttcacaacacatatccagaggggcctgca 4436
          || || ||| |  | || ||   | | ||   |||
Db   1555 TACGAGTTCTCGTCGCACATAACGCGAGGGGAGCA 1589


RESULT 13
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             0.6%;  Score 57.4;  DB 1;  Length 5852;
  Best Local Similarity   46.1%;  Pred. No. 0.0003;
  Matches  193;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy   7160 acaaatagttctttatttttggcatgtatgccttttttattttcattcaatttttttttt 7219
          | ||| |  |    || || |      | | || | || ||||  ||   |||| |  |
Db   1590 ATAAAAAACTAACCTAGTTATAATTACTTTCCCCTCTTTTTTTTTTTTTTTTTGTCATG 1649

Qy   7220 ttctcagacagacatagtagtatcaactagcattggaaaatacatatcactattcttgga 7279
                    | ||  |  ||  |    ||  |||| | | |  |  ||  |   |
Db   1650 ACACTTTTTTTTTTTTGTCATGACACTTTTTTTTTAAAAAAAAAAAAAAAAAATGTTAAAA 1709

Qy   7280 atatttatggtcagtctacttttttagtaaaatatttttggatagcgttgacacgatagat 7339
             | | || | |   |  |||||  ||   | ||||| |||||   | | |  | |
Db   1710 TACTATTTGATGACATTCATTTTTCCTAGTTTTTTTTTAGATAGATATAAAAATAAATTG 1769

Qy   7340 cttattccatacttctttattattgataatttttattttcatttttttgctttcattattat 7399
          | |||        | ||| |||  | |  |||  ||  |  |||  |  ||      ||| |
Db   1770 CCTATCGATATATACTTAATTTATTAAGATTGAATAATATTTTAATTTTTAATAAATTCT 1829

Qy   7400 acatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataaca 7459
             || | ||||  |           ||        ||| |||           ||||||| | |
Db   1830 ACTTTTTTTTTTTTTTTTCTTTTTTTTTTTAAATTTTAAAATTTTTTTTTTTTTTATTAGATCT 1889

Qy   7460 ctaaacactccttttttgacatattaaagcctttattccatctctcaagatatattataa 7519
          |  ||        |   |   | ||||||   | | ||  |     ||| |  ||| | ||
Db   1890 CATAATTAAAAATCAATTTAAAATTAAAAGTTATTTTTAAATATGCAAAAACTATAAAAA 1949

Qy   7520 aatttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaac 7578
          | |    |  |||||   |  |||   | || |||| | ||   || |  ||||  ||
Db   1950 ACTAATGTAGTTTAACCAACTTTTTTCTATTTCTTTTTTTTTTTTTTTTTTTTTTTAC 2008


RESULT 14
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
```

```
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             0.6%;  Score 56.8;  DB 2;  Length 19124;
  Best Local Similarity   45.6%;  Pred. No. 0.00083;
  Matches  248;  Conservative    0;  Mismatches  287;  Indels    9;  Gaps    1;

Qy   7193 ttttattttcattcaatttttttttttttctcagacagacatagtagtatcaactagcat 7252
          ||||   ||| ||  |  | ||| | |   | | | |   |||   || |  |  |  ||
Db    680 TTTTCGATTTTATCGATATATTTATGTCGTTTATATACTTATATAGGTCTTTATAACTAT 739

Qy   7253 tggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaata 7312
          ||   |||| |    | ||  ||   |||||  ||  ||    |  |   |      |||
Db    740 TGATTAATAGAAGGTAATAGCCTAATAATATAAATACTCGTATTTATAAATTCATTTATA 799

Qy   7313 tttttggatagcgttgacacgatagatcttattccatacttctttattattgataatttt 7372
          | |||  |    ||   | | |  || ||     | | ||  |||   ||  ||| | |
Db    800 TATTTCAAATATATTTCGATGGTTTATTTTCAAATACAATTAATTAGATTTCTTAAATAT 859

Qy   7373 attttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttt 7432
          | |||||||  ||  |||   ||  ||| | ||         | | |   |    | ||
Db    860 TTCTTCATTTATTCATTTTTATAGCATATACAT---------GCACATTATAAATTATTA 910

Qy   7433 tgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagcctt 7492
          |||  |      | |||| ||||  |  | |   | | |   | |||| ||  |  |
Db    911 ATAAAAAATTTTTATTTTAATATATAATAACAATTTTCATACATTACATTTTTCACACAA 970

Qy   7493 tattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatta 7552
          |||  |  | |||  || ||  | |  |  || | | ||| ||  | || | | | |
Db    971 CATTTAAGTTGTCATAATGTAACACATTAAATAATATATTACTTATATATATATAATTAT 1030

Qy   7553 ttttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgagg 7612
          |  |    ||     | |   |  | | |   | |  ||    | | | | |||||  |
Db   1031 TAATTATATATTAAATAAAAATGTATTATCGCCTGTATTATCATAGTATATATAATGTTG 1090

Qy   7613 aatttcatgatgatttaggaatttttctctcttggaaaaggcttcccctgtgatgaaaatg 7672
          ||  |      |  ||   ||  |          ||||   |     |   |   || |
Db   1091 TATAACGCTTCAAAATATATATAATAATATAATTAAAAATATATATATAGTAATTAATTA 1150

Qy   7673 atgtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtct 7732
          | ||  |  | |  |  |    || |||| |   || || ||| || |  ||||  |
Db   1151 TTTTGTTATGTTATGTAATAATGCAATTAATATAAGATAAAATTCTATAGCTATTATTTA 1210

Qy   7733 atat 7736
          | ||
Db   1211 AAAT 1214


RESULT  15
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             0.5%;  Score 54;  DB 1;  Length 7218;
  Best Local Similarity   2.1%;  Pred. No. 0.0024;
  Matches    6;  Conservative  181;  Mismatches  101;  Indels    0;  Gaps    0;

Qy   8551 tttggtttgagagccatggcaaaaaaaaaaaaagaaaaaaagatcaagtcgtctttggtg 8610
          ||||||  :::::  : :: :::::::::::::::::::::::::  :::  :     :: :
Db   1442 TTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1383

Qy   8611 agccagtaaggtgaaagcttgctgactgtccaaggcacaagagaaaattgaggaattgaa 8670
```

```
          ::  :: :::: :::::   :  ::  :   :::: : :::::::::  ::::::  :::
Db   1382 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1323

Qy   8671 atgcaacctgagtatcaaactaaatattctaatcaaaggtaggtactgttaggtggaatt 8730
          : : ::   ::: :  :::  ::: :    ::  ::::: ::: :  :  ::: ::::
Db   1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1263

Qy   8731 ctatcagcaggcaactgcaaatgagaagaagatagaaggacgcccgtcgggactttggag 8790
            :  :: ::: ::  : ::: :::::::::: :::::::: :   :  ::::    ::::
Db   1262 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1203

Qy   8791 ggcattgttattttcccaaagaaagacggccaagggcagaggcatgga 8838
          :: :  :  :       ::::::::: ::  ::::: ::::: : :::
Db   1202 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155
```

Search completed: May 1, 2001, 00:38:40
Job time: 56864 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:18:01 ; Search time 12176.5 Seconds
(without alignments)
120.109 Million cell updates/sec

Title: US-09-721-183-20
Perfect score: 10006
Sequence: 1 ttcctccgcgaaggctcctt.....agttttccaccagttacaca 10006

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	837.2	8.4	842	106	AL535097	AL535097 AL535097
	2	701.4	7.0	824	174	BG254038	BG254038 602366975
	3	690	6.9	783	31	AV687506	AV687506 AV687506
	4	681.6	6.8	691	107	AU117444	AU117444 AU117444
	5	664.4	6.6	742	107	AU117220	AU117220 AU117220
	6	649.2	6.5	1043	171	BF982937	BF982937 602306330
	7	648.6	6.5	813	107	AU117784	AU117784 AU117784
	8	630	6.3	649	107	AU117175	AU117175 AU117175
c	9	594.8	5.9	598	113	AW303579	AW303579 xv20g11.x
	10	592.2	5.9	964	173	BG115118	BG115118 602315968
	11	587.6	5.9	917	174	BG178176	BG178176 602327784
	12	587.2	5.9	953	144	BF138096	BF138096 601785006
c	13	571.8	5.7	602	229	AG025673	AG025673 Homo sapi
	14	567.8	5.7	749	144	BF139707	BF139707 601785355
	15	557	5.6	569	32	AW021592	AW021592 df25f07.y
c	16	547	5.5	547	103	AI858596	AI858596 wl40c11.x
c	17	546.4	5.5	584	122	AW973691	AW973691 EST385791
c	18	541.6	5.4	569	107	AU144074	AU144074 AU144074

	19	539.8	5.4	567	141	BE932151	BE932151 IL3-HT061
	20	531	5.3	729	144	BF168531	BF168531 601775431
c	21	529.4	5.3	543	22	AI590321	AI590321 tn49a11.x
c	22	529	5.3	542	104	AI953076	AI953076 wp99a05.x
c	23	525	5.2	525	148	BF439110	BF439110 nab59d09.
c	24	520.8	5.2	560	107	AU144274	AU144274 AU144274
	25	520.2	5.2	566	164	BE234097	BE234097 140704 MA
c	26	513.6	5.1	521	10	AA631027	AA631027 nq76h02.s
	27	505.6	5.1	847	144	BF139728	BF139728 601785379
c	28	505	5.0	528	7	AA428090	AA428090 zw32a08.r
c	29	503	5.0	503	110	AW070915	AW070915 xa31c12.x
	30	498.8	5.0	519	168	BF744693	BF744693 QV2-BT061
c	31	498.8	5.0	571	107	AU144115	AU144115 AU144115
	32	498.4	5.0	667	174	BG174562	BG174562 602334405
c	33	492	4.9	562	168	BF744704	BF744704 QV2-BT061
	34	486.2	4.9	925	144	BF160771	BF160771 601769175
	35	483.6	4.8	569	165	BE311148	BE311148 601057836
c	36	478	4.8	478	19	AI380343	AI380343 tf94h10.x
c	37	474.6	4.7	572	107	AU144524	AU144524 AU144524
c	38	473.6	4.7	509	168	BF747629	BF747629 QV2-BT061
c	39	472	4.7	542	168	BF747621	BF747621 QV2-BT061
	40	472	4.7	664	165	BE283806	BE283806 601104087
	41	469.4	4.7	693	174	BG246039	BG246039 602358869
c	42	461	4.6	461	22	AI610247	AI610247 tp14a12.x
	43	459.4	4.6	481	111	AW149608	AW149608 xf40b02.x
c	44	459	4.6	467	103	AI927871	AI927871 wo66c11.x
	45	457.8	4.6	640	141	BE912092	BE912092 601663271

ALIGNMENTS

```
RESULT   1
AL535097
LOCUS       AL535097      842 bp    mRNA            EST       13-FEB-2001
DEFINITION  AL535097 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF008YI01 5
            prime, mRNA sequence.
ACCESSION   AL535097
VERSION     AL535097.1  GI:12798590
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DF008YI01"
                     /clone_lib="LTI_FL013_FBrn1"
                     /dev_stage="pooled tissue from post conception fetuses (20
                     week, 24 week and 26 week)"
                     /lab_host="DH10B"
                     /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector.  Library was constructed by Life
                     Technologies.  Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      249 a    130 c    126 g    337 t
ORIGIN
```

```
  Query Match             8.4%;  Score 837.2;  DB 106;  Length 842;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-170;
  Matches  839;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    7087 aacttgggcaatatgacaagtattacttttttgggcaaaactactcattaagcaatttctc 7146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATTAAGCAATTTCTC 60

Qy    7147 tagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttttatttttcatt 7206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTTTTTTATTTTCATT 120

Qy    7207 caatttttttttttctcagacagacatagtagtatcaactagcattggaaaatacatat 7266
           |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db     121 CAATTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATAT 180

Qy    7267 cactattcttggaatatttatggtcagtctacttttttagtaaaatatttttggatagcgt 7326
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     181 CACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAGAATATTTTTGGATAGCGT 240

Qy    7327 tgacacgatagatcttattccatacttctttattattgataattttattttcatttttttg 7386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTG 300

Qy    7387 ctttcattattatacatattttggtggagaagaggttgggctttttttgaaagagacaaaa 7446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAA 360

Qy    7447 atttattataacactaaacactccttttttgacatattaaagcctttattccatctctca 7506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCA 420

Qy    7507 agatatattataaaatttatttttttaatttaagatttctgaattatttttatcttaaatt 7566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATT 480

Qy    7567 gtgattttaaacgagctattatggtacggaacttttttttaatgaggaatttcatgatgat 7626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTTAATGAGGAATTTCATGATGAT 540

Qy    7627 ttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaa 7686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTGCCAGCTAAA 600

Qy    7687 attgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatattctaaattga 7746
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATTCTAAATTGA 660

Qy    7747 gctttggatcaaactttaggccaggaccagctcatgcgttctcattcttccttttctcac 7806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCCTTTTCTCAC 720

Qy    7807 tctttctctcatcactcacctctgtattcattctgttgtttgggatagaaaaatcataaa 7866
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAAAATCATAAA 780

Qy    7867 gagccaacccatctcagaacgttgtggattgagagagacactacatgactccaagtatat 7926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTCCAAGTATAT 840

Qy    7927 ga 7928
           ||
Db     841 GA 842
```

```
RESULT   2
BG254038
LOCUS       BG254038      824 bp    mRNA            EST       13-FEB-2001
DEFINITION  602366975F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4475131 5',
            mRNA sequence.
ACCESSION   BG254038
```

```
VERSION     BG254038.1  GI:12763844
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10300  row: e  column: 20
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4475131"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      265 a    205 c    180 g    174 t
ORIGIN
```

Query Match 7.0%; Score 701.4; DB 174; Length 824;
Best Local Similarity 95.0%; Pred. No. 3.1e-141;
Matches 767; Conservative 0; Mismatches 36; Indels 4; Gaps 4;

```
Qy   3396 gtggcctctaccagaagcttcactcgactcccaggccttaaacatcattaaacaaaaca 3455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GTGGCCTCTACCAGAAGCTTCACTCGACTCCCAGGCCTTTAAACATCATTAAACAAAACA 65

Qy   3456 acggtgagcagattattaggaggagaacaagaaagcgccttaacccagaggcacttcagg 3515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 ACGGTGAGCAGATTATTAGGAGGAGAACAAGAAAGCGCCTTAACCCAGAGGCACTTCAGG 125

Qy   3516 ctgagcagctcaacaaacagcagaggggcagcaatgaggagcaagtcaatggaagcccgt 3575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 CTGAGCAGCTCAACAAACAGCAGAGGGGCAGCAATGAGGAGCAAGTCAATGGAAGCCCGT 185

Qy   3576 tagagaggaggtcagaagatcatctaactgaaagtcaccagagagaaattccactcccca 3635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 TAGAGAGGAGGTCAGAAGATCATCTAACTGAAAGTCACCAGAGAGAAATTCCACTCCCCA 245

Qy   3636 gcctaagtaaatacgaagcccagggttcattgactaaaagccattctgctcagcagccag 3695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 GCCTAAGTAAATACGAAGCCCAGGGTTCATTGACTAAAAGCCATTCTGCTCAGCAGCCAG 305

Qy   3696 tcctggtcagccaaactctggatattcacaaaaggatgcaacctttgcacattcagataa 3755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 TCCTGGTCAGCCAAACTCTGGATATTCACAAAAGGATGCAACCTTTGCACATTCAGATAA 365

Qy   3756 aaagtcctcaggaaagtactggagatccaggaaatagttcatccgtatctgaagggaaag 3815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 AAAGTCCTCAGGAAAGTACTGGAGATCCAGGAAATAGTTCATCCGTATCTGAAGGGAAAG 425

Qy   3816 gaagttctgagagaggcagtcctatagaaaagtacatgagacctgcgaaacacccaaatt 3875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GAAGTTCTGAGAGAGGCAGTCCTATAGAAAAGTACATGAGACCTGCGAAACACCCAAATT 485

Qy   3876 attcaccaccaggcagccctattgaaaagtaccagtacccacttttggacttccctttt- 3934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    486 ATTCACCACCAGGCAGCCCTATTGAAAAGTACCAGTACCCACTTTTTGGACTTCCCTTTG 545

Qy   3935 tacataatgacttccagagtgaagctgattggctgcggttctggagtaaatataagctct 3994
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    546 TACATAATGACTTCCAGAGTGAAGCTGATTGGCTGCGGTTCTGGAGT-AATATAAGCTCT 604

Qy   3995 ccgttcct-gggaatccgcactacttgagtcacgtgcctggcctaccaaatccttgccaa 4053
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 CCGTTCCTGGGGAATCCGCACTACTTGAGTCACGTGCCTGGCCTACCAAATCCTTGCCAA 664

Qy   4054 aactatgtgccttatcccaccttcaatctgcctcctcatttttcagctgttggatcagac 4113
          |||||||||||||||| | |||||||||| |  ||||| ||||||||||||||
Db    665 AACTATGTGCCTTATCCACCTTTCAATCTGCTCCTCAATTTTCCAGCTGTTGGATCAAGA 724

Qy   4114 aatgacattcctctagatttggcgatcaagcattccagacctgggccaactgcaaacggt 4173
          |    | | |      | | | |||||||||| ||||| ||||| ||||||||||||||
Db    725 CAATGAAATTCCTCCTAGATTGGGATCAAGCATACCAGAACTGGG-CAACTGCAAACGGT 783

Qy   4174 gcctccaaggagaaaacgaaggcacca 4200
          ||||||||||||||   |   |||||
Db    784 GCCTCCAAGGAGAAACGAAGCACACCA 810
```

```
RESULT   3
AV687506
LOCUS       AV687506      783 bp    mRNA            EST       25-SEP-2000
DEFINITION  AV687506 GKC Homo sapiens cDNA clone GKCDRB04 5', mRNA sequence.
ACCESSION   AV687506
VERSION     AV687506.1  GI:10289369
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 783)
  AUTHORS   Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            ,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA GK- clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .783
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKCDRB04"
                     /clone_lib="GKC"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      256 a    143 c    152 g    230 t      2 others
ORIGIN
```

Query Match 6.9%; Score 690; DB 31; Length 783;
Best Local Similarity 95.2%; Pred. No. 9.1e-139;
Matches 756; Conservative 0; Mismatches 27; Indels 11; Gaps 4;

```
Qy   4443 caatgcacaagtggaaaaaaatggaaaacctaaagagtaaaaccttagcacttagcacaa 4502
```

```
Db      1 CAATGCACAGGGGGAGAAAAATGGAAAACCTAAAGAGTAAAACCTTAGCACTTAGCACAA 60

Qy   4503 ttaaatagaaataggttttcttgatgggaattcaatagcttgtaatgtcttatgaagacc 4562
Db     61 TTAAATAGAAATAGGTTGTCTTGATGGGAATTCAATAGCTTGTAATGTCTTATGAAGACC 120

Qy   4563 tattaaaaaaaatacttcatagagcctgccttatccaacatgaaattcccttcttttgtta 4622
Db    121 TATTAAAAAAATACTTCATAGAGCCTGCCTTATCCAACATGAAATTCCCTTCTTTTGTTA 180

Qy   4623 ttctttcttttgatgagtaggttaccaagattaaaaagtgagataaatggtcaatgagaa 4682
Db    181 TTCTTTCTTTTGATGAGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGAGAA 240

Qy   4683 agaatggaagatggtaaacaatcacttttaaaacctgttaagtcaaaaccatcttggct 4742
Db    241 AGAATGGAAGATGGTGAACAATCACTCTTTAAAACCTGTTAAGTCAAAACCATCTTGGCT 300

Qy   4743 aatatgtactggggaaataatccataagagatatcaccagactagaattaatatatttat 4802
Db    301 AATATGTACTGGGGAAATAATCCATATGAGATATCACCAGACTAGAATTAATATATTTAT 360

Qy   4803 aaagaaagagaccaaaactgtctagaatttgaaagggtttacatattattatactaaagc 4862
Db    361 AAAGAAAGAGACCAAAACTGTCTAGAATTTGAAAGGGGTTACATATTATTATACTAAAGC 420

Qy   4863 agtactggactggccattggaccatttgttccaaaacccataaattgttgcctaaattta 4922
Db    421 AGTACTGGACCGGCCATTGGACCATATGTTCCAAAACCCATAAATTGTTGCCTAAATTTA 480

Qy   4923 taatgatcatgaaaccctaggcagaggaggagaaattgaaggtccagggcaatgaaagaa 4982
Db    481 TAATGATCATGAAACCCTATGCAGAGGAGGAGAAATTGAAGGTCCACGGCAATGAAAG-- 538

Qy   4983 aaatggcgccctctcaatttagtcttctctcattggccatgtttcagattttgacctaga 5042
Db    539 --ATGGCGCCCTCTCAATTTAGACTTCTCTCATTGGCCATGTGTCAGATTTTGACCT--- 593

Qy   5043 aatgcgagctgtggttaggcttggttagagtgcagcaagcaacatgacagatggtggcac 5102
Db    594 ATTGCGAGCTGTGGTTAGGCTTGGTTAGAGTGCAGCAAGCAACATGACAGATGGTGGCAC 653

Qy   5103 gctgtttttacccagccctgcctgtacatacacatgcacaccctctctgatatttttgtc 5162
Db    654 GCCTGTTTTACCCAGACCTGCCTGTACATACACATGCACACCCTCTCTGATGATTTTTGC 713

Qy   5163 ctttagatgttcaaatactcagtagtccttttgtttgcggtttagattcattttgtccac 5222
Db    714 CTTTAGATGTTC---TACTCAGTAGTNCCTCTGTTTGCGGNTTAGATTCATTTTGT-CAC 769

Qy   5223 acatgtacccattt 5236
Db    770 ACATGTACCCATTT 783
```

```
RESULT   4
AU117444
LOCUS       AU117444      691 bp    mRNA            EST       19-OCT-2000
DEFINITION  AU117444 HEMBA1 Homo sapiens cDNA clone HEMBA1001411 5', mRNA
            sequence.
ACCESSION   AU117444
VERSION     AU117444.1  GI:10932413
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1001411"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      211 a     93 c    104 g    280 t      3 others
ORIGIN
```

```
  Query Match             6.8%;  Score 681.6;  DB 107;  Length 691;
  Best Local Similarity   99.0%;  Pred. No. 5.9e-137;
  Matches  684;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy   6985 taatttttcttagtaaaaagaataacagaatgcatcgtggcaatccttaagcaacattatc 7044
Db      1 TAATTTTCTTAGTAAAAAGAATAACAGAATGCATCGTGGCAATCCTTAAGCAACATTATC 60

Qy   7045 tatgtggactgcttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgaca 7104
Db     61 TATGTGGACTGCTTAAATCAGCAAAACACCAGAAGTTTGGTTAACTTGGGCAATATGACA 120

Qy   7105 agtattacttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaa 7164
Db    121 AGTATTACTTTTTGGGCAAAACTACTCATTAAGCAATTTCTCTAGTGTGTCGGACACAAA 180

Qy   7165 taggttctttatttttggcatgtatgccttttttatttttcattcaatttttttttttctc 7224
Db    181 TAGGTTCTTTATTTTTGGCATGTATGCCTTTTTATTTTCATTCAATTTTTTTTTTTTCTC 240

Qy   7225 agacagacatagtagtatcaactagcattggaaaatacatatcactattcttggaatatt 7284
Db    241 AGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATATCACTATTCTTGGAATATT 300

Qy   7285 tatggtcagtctacttttttagtagaaatatttttggatagcgttgacacgatagatcttat 7344
Db    301 TATGGTCAGTCTACTTTTTTAGTAGAATATTTTTGGATAGCGTTGACACGATAGATCTTAT 360

Qy   7345 tccatacttctttattattgataatttttattttcatttttgctttcattattatacata 7404
Db    361 TCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTGCTTTCATTATTATACATA 420

Qy   7405 ttttggtggagaagaggttgggctttttttgaaagagacaaaaatttattatataacactaaa 7464
Db    421 TTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAAATTTATTATAACACTAAA 480

Qy   7465 cactccttttttgacatattaaagcctttattccatctctcaagatatattataaaaattt 7524
Db    481 CACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCAAGATATATTATAAAATTT 540

Qy   7525 attttttttaatttaagatttctgaattatttttatcttaaattgtgattttaaacgagcta 7584
Db    541 ATTTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATTGTGATTTTAAGCGAGCTA 600

Qy   7585 ttatggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttctctctt 7644
Db    601 TTATGGTACGGAACTTTTTTTAATGAGGAATTTCATGATGATTTANGAATTTTCTCTCTT 660
```

```
Qy   7645 ggaaaaggcttcccctgtgatgaaaatgatg 7675
          ||||||||||| ||||| ||||||||||||||
Db    661 GGAAAAGGCTTNCCCTGNGATGAAAATGATG 691


RESULT   5
AU117220
LOCUS       AU117220      742 bp    mRNA            EST       19-OCT-2000
DEFINITION  AU117220 HEMBA1 Homo sapiens cDNA clone HEMBA1000916 5', mRNA
            sequence.
ACCESSION   AU117220
VERSION     AU117220.1  GI:10932168
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1000916"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      229 a    183 c    161 g    164 t      5 others
ORIGIN


  Query Match           6.6%;  Score 664.4;  DB 107;  Length 742;
  Best Local Similarity  98.1%;  Pred. No. 3.2e-133;
  Matches  712;  Conservative    0;  Mismatches   10;  Indels    4;  Gaps    4;

Qy   3385 atgcaacgcgtgtggcctctacc-agaagcttcactcgactcccaggcctttaaacatca 3443
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      1 ATGCAACGCGTGTGGCCTCTACCAAGAAGCTTCACTCGACTCCCAGGCCTTTAAACATCA 60

Qy   3444 ttaaacaaaacaacggtgagcagattattaggaggagaacaagaaagcgccttaacccag 3503
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     61 TTAAACAAAACAACGGTGAGCAGGTTATTAGGAGGAGAACAAGAAAGCGCCTTAACCCAG 120

Qy   3504 aggcacttcaggctgagcagctcaacaaacagcagaggggcagcaatgaggagcaagtca 3563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAGGGGCAGCAATGAGGAGCAAGTCA 180

Qy   3564 atggaagcccgttagagaggaggtcagaagatcatctaactgaaagtcaccagagagaaa 3623
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCTAACTGAAAGTCACCAGAGAGAAA 240

Qy   3624 ttccactccccagcctaagtaaatacgaagcccagggttcattgactaaaagccattctg 3683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCCACTCCCCAGCCTAAGTAAATACGAAGCCCAGGGTTCATTGACTAAAAGCCATTCTG 300

Qy   3684 ctcagcagccagtcctggtcagccaaactctggatattcacaaaaggatgcaacctttgc 3743
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTCAGCAGCCAGTCCTGGTCAGCCAAACTCTGGATATTCACAAAAGGATGCAACCTTTGC 360

Qy   3744 acattcagataaaaagtcctcaggaaagtactggagatccaggaaatagttcatccgtat 3803
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACATTCAGATAAAAAGTCCTCAGGAAAGTACTGGAGATCCAGGAAATAGTTCATCCGTAT 420

Qy   3804 ctgaagggaaaggaagttctgagagaggcagtcctatagaaaagtacatgagacctgcga 3863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGAAGGGAAAGGAAGTTCTGAGAGAGGCAGTCCTATAGAAAAGTACATGAGACCTGCGA 480

Qy   3864 aacacccaaattattcaccaccaggcagccctattgaaaagtaccagtacccacttttg 3923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACACCCAAATTATTCACCACCAGGCAGCCCTATTGAAAAGTACCAGTACCCACTTTTG 540

Qy   3924 gacttccctttttacataatgacttccagagtgaagctga-ttggctgcggttctggagta 3982
          | |||||||||  ||||||||||||||||||||||| |||| ||||||||||||||| |||
Db    541 GNCTTCCCTTTGTCATAATGACTTCCAGAGTGAANCTGATTTGGCTGCGGTTCTGGGGTA 600

Qy   3983 aatat-aagctctccgttcctgggaatccgcactacttgagtcacgtgcctggcctacca 4041
          |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATATAAAGCTCTCCGTTCCTGGGAATCCGCACTACTTGAGTCACGTGCCTGGCCTACCA 660

Qy   4042 aatccttgccaaaactatgtgccttatcccaccttc-aatctgcctcctcatttttcagc 4100
          |||||||||||||||||||||||||||||| |||| |||||||||||||| |||||| |
Db    661 AATCCTTGCCAAAACTATGTGCCTTATCCCANCTTCAAATCTGCCTCCTCANTTTTCAAC 720

Qy   4101 tgttgg 4106
          |||| |
Db    721 TGTTTG 726


RESULT   6
BF982937
LOCUS       BF982937      1043 bp    mRNA            EST       23-JAN-2001
DEFINITION  602306330F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4397719 5',
            mRNA sequence.
ACCESSION   BF982937
VERSION     BF982937.1  GI:12385749
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10098  row: l  column: 08
            High quality sequence stop: 642.
FEATURES             Location/Qualifiers
     source          1. .1043
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4397719"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      317 a    260 c    293 g    173 t
```

```
ORIGIN

  Query Match             6.5%;  Score 649.2;  DB 171;  Length 1043;
  Best Local Similarity   95.1%;  Pred. No. 6.5e-130;
  Matches  702;  Conservative    0;  Mismatches   33;  Indels    3;  Gaps    3;

Qy   3021 gaggggcagacatcctgcgggggagtccgtcatacacccaagcaagcctggggctgctga 3080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 GAGGGGCAGACATCCTGCGGGGGAGTCCGTCATACACCCAAGCAAGCCTGGGGCTGCTGA 70

Qy   3081 cgcctgtgtctggcacccaagagcagac-aaagactctaagggatagtcccaatgtggag 3139
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     71 CGCCTGTGTCTGGCACCCAAGAGCAGACGAAAGACTCTAAGGGATAGTCCCAATGTGGAG 130

Qy   3140 gccgcccatctggcgcgacctatttatggcttggctgtggaaaccaagggattcctgcag 3199
          |||  |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    131 GCCAGCCATCTGGCGCGAGCTATTTATGGCTTGGCTGTGGAAACCAAGGGATTCCTGCAG 190

Qy   3200 ggggcgccagctggcggagagaagtctggggccctcccccagcagtatcctgcatcggga 3259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 GGGGCGCCAGCTGGCGGAGAGAAGTCTGGGGCCCTCCCCCAGCAGTATCCTGCATCGGGA 250

Qy   3260 gaaaacaagtccaaggatgaatcccagtccctgttacggaggcgtagaggctccggtgtt 3319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 GAAAACAAGTCCAAGGATGAATCCCAGTCCCTGTTACGGAGGCGTAGAGGCTCCGGTGTT 310

Qy   3320 ttttgtgccaattgcctgaccacaaagacctctctctggcgaaagaatgcaaatggcgga 3379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TTTTGTGCCAATTGCCTGACCACAAAGACCTCTCTCTGGCGAAAGAATGCAAATGGCGGA 370

Qy   3380 tatgtatgcaacgcgtgtggcctctaccagaagcttcactcgactcccaggcctttaaac 3439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 TATGTATGCAACGCGTGTGGCCTCTACCAGAAGCTTCACTCGACTCCCAGGCCTTTAAAC 430

Qy   3440 atcattaaacaaaacaacggtgagcagattattaggaggagaacaagaaagcgccttaac 3499
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    431 ATCATTAAACAAAACAACGGTGAGCAGATTATTAGGAGGAGAACGAGAAAGCGCCTTAAC 490

Qy   3500 ccagaggcacttcaggctgagcagctcaacaaacagcagaggggcagcaatgaggagcaa 3559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 CCAGAGGCACTTCAGGCTGAGCAGCTCAACAAACAGCAGAGGGGCAGCAATGAGGAGCAA 550

Qy   3560 gtcaatggaagcccgttagagaggaggtcagaagatcatctaactgaaagtcaccagaga 3619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 GTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAGATCATCTAACTGAAAGTCACCAGAGA 610

Qy   3620 gaaattccactccccagcctaagtaaatacgaagcccagggtt-cattgactaaaagcca 3678
          ||||||| | ||||||| |||||||||||||||| |||||||| |||||| |||||| ||
Db    611 GAAATTCGA-TCCCCAGACTAAGTAAATACGAAGGCCAGGGTTCCATTGAGTAAAAGGCA 669

Qy   3679 ttctgctcagcagccagtcctggtcagccaaactctggatattcacaaaaggatgcaacc 3738
          ||||| ||||||| |||||  | ||||| |||||||||   |||   ||||| |||| ||
Db    670 TTCTGGTCAGCAGGCAGTCTGGTTCAGCAAAACTCTGGGATTTCCACAAAGGTTGCACCC 729

Qy   3739 tttgcacattcagataaa 3756
          ||    | |  | |  ||
Db    730 TTGCGCCTTAGATAACAA 747


RESULT   7
AU117784
LOCUS       AU117784      813 bp    mRNA            EST       19-OCT-2000
DEFINITION  AU117784 HEMBA1 Homo sapiens cDNA clone HEMBA1002177 5', mRNA
            sequence.
ACCESSION   AU117784
VERSION     AU117784.1  GI:10932779
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1002177"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      250 a    193 c    183 g    184 t      3 others
ORIGIN

  Query Match             6.5%;  Score 648.6;  DB 107;  Length 813;
  Best Local Similarity   96.3%;  Pred. No. 8.5e-130;
  Matches  685;  Conservative    0;  Mismatches   21;  Indels    5;  Gaps    2;

Qy   2106 cagaaggagagacaatgaccaagacagacaagagctcgagtggggctaaaaagaaggact 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 CAGAAGGAGAGACAATGACCAAGACAGACAAGAGCTCGAGTGGGGCTAAAAAGAAGGACT 156

Qy   2166 tctccagcaagggagccgaggataatatggtaacgagctataattgtcagttctgtgact 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 TCTCCAGCAAGGGAGCCGAGGATAATATGGTAACGAGCTATAATTGTCAGTTCTGTGACT 216

Qy   2226 tccgatattccaaaagccatggccctgatgtaattgtagtggggccacttctccgtcatt 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 TCCGATATTCCAAAAGCCATGGCCCTGATGTAATTGTAGTGGGGCCACTTCTCCGTCATT 276

Qy   2286 atcaacagctccataacattcacaagtgtaccattaaacactgtccattctgtcccagag 2345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 ATCAACAGCTCCATAACATTCACAAGTGTACCATTAAACACTGTCCATTCTGTCCCAGAG 336

Qy   2346 gactttgcagcccagaaaagcaccttggagaaattacttatccgtttgcttgtagaaaaa 2405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 GACTTTGCAGCCCAGAAAAGCACCTTGGAGAAATTACTTATCCGTTTGCTTGTAGAAAAA 396

Qy   2406 gtaattgttcccactgtgcactcttgcttctgcacttgtctcctggggcggctggaagct 2465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 GTAATTGTTCCCACTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGCGGCTGGAAGCT 456

Qy   2466 cgcgagtcaaacatcagtgccatcagtgttcattcaccacccctgacgtagatgtactcc 2525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CGCGAGTCAAACATCAGTGCCATCAGTGTTCATTCACCACCCCTGACGTAGATGTACTCC 516

Qy   2526 tctttcactatgaaagtgtgcatgagtcccaagcatcggatgtcaaacaagaagcaaatc 2585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 TCTTTCACTATGAAAGTGTGCATGAGTCCCAAGCATCGGATGTCAAACAAGAAGCAAATC 576

Qy   2586 acctgcaaggatcggatgggcagcagtctgtcaaggaaagcaaagaacactcatgtacca 2645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 ACCTGCAAGGATCGGATGGGCAGCAGTCTGTCAAGGAAAGCAAAGAACACTCATGTACCA 636

Qy   2646 aatgtgattttattacccaagtggaagaagagatttcccgacactacaggagagcacaca 2705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 AATGTGATTTTATTACCCAAGTGGAAGAAGAGATTTCCCGACACTACAGGAGAGCACACA 696
```

```
Qy   2706 gctgctacaaatgccgtcagtgcagttttacagct--gccgatactcagtcactactgga 2763
          ||||||||||||||||||||||||||||||||||    |   |  ||||||||||||||
Db    697 GCTGCTACAAATGCCGTCAGTGCAGTTTTACAGCTTGCCGATTCTTCAGTCACTACTGGA 756

Qy   2764 gcacttcaacactgtt---cactgccaggaacaggacatcactacagccaa 2811
          |||||| ||||||  |   |    |  | | |||||||| ||| ||| | |
Db    757 GCACTTTAACACTTGTTCACTTGCCAGGAACCAGGACATNACTNCAGGCCA 807


RESULT   8
AU117175
LOCUS       AU117175      649 bp    mRNA            EST       19-OCT-2000
DEFINITION  AU117175 HEMBA1 Homo sapiens cDNA clone HEMBA1000822 5', mRNA
            sequence.
ACCESSION   AU117175
VERSION     AU117175.1  GI:10932121
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1000822"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      207 a    109 c    121 g    209 t      3 others
ORIGIN


  Query Match             6.3%;  Score 630;  DB 107;  Length 649;
  Best Local Similarity   99.5%;  Pred. No. 8.7e-126;
  Matches  641;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy   5304 gatgaaccagtatggatttgctttttctaagcctcctgttggttactaatctcacttggca 5363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATGAACCAGTATGGATTTGCTTTTCTAAGCCTCCTGTTGGTTACTAATCTCACTTGGCA 60

Qy   5364 cattataactaaaggaatcccctcaattcaaaagcatagatggatacaaatgtcagaccg 5423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CATTATAACTAAAGGAATCCCCTCAATTCAAAAGCATAGATGGATACAAATGTCAGACCG 120

Qy   5424 tgggtttaatttgtttagaacacatggcatttcttcacaaggtaacctgctgtatttatt 5483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGTTTAATTTGTTTAGAACACATGGCATTTCTTCACAAGGTAACCTGCTGTATTTATT 180

Qy   5484 tattttcttttggttaaatataatttccaaactttgtggtcaggcagcgtctaaggttac 5543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATTTTCTTTTGGTTAAATATAATTTCCAAACTTTGTGGTCAGGCAGCGTCTAAGGTTAC 240

Qy   5544 gttaccacagactgacagttggtatatgtaccagccaatcccttcattaaatgtatacag 5603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTACCACAGACTGACAGTTGGTATATGTACCAGCCAATCCCTTCATTAAATGTATACAG 300

Qy   5604 atttagttaagtagcattaaataggattcttagaagtatgtcctcatagaacttttaata 5663
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAGTATGTCCTCATAGAACTTTTAATA 360

Qy   5664 cttaaggctttgtaaaaactatccatgaagggaaagctcctcagcataactgctcaggga 5723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAGCTCCTCAGCATAACTGCTCAGGGA 420

Qy   5724 aatagggctaaataactgaacattaaataattggttaaaggtgctgttagtcgagcctca 5783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATAGGGCTAAATAACTGAACATTAAATAATTGGTTAAAGGTGCTGTTAGTCGAGCCTCA 480

Qy   5784 atgcttgctacaaggatgtatgtacaaggactgactttaataatttgcattatattgtcc 5843
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGCTTGCTACAAGGATGTATGTACAAGGACTGACTTTAATAATTTGCATTATATTGTCC 540

Qy   5844 caaccagtagtttatttttgccacggagatgtagaagatattacaagctactggatgca 5903
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    541 CAACCAGTAGTTTATTTTTGCCACGGAGATGTANAAGATATTACAAGCTACTGGATGCA 600

Qy   5904 ct-gtcagattaacttatttcattaaagaagttgggagaacaaa 5946
          || |||||||||||| |||||||||||||||||||||||||||||
Db    601 CTGGTCAGATTAACNTATTTCATTAAAGAAGTTGGGAGAACAAA 644


RESULT   9
AW303579/c
LOCUS       AW303579      598 bp    mRNA            EST       18-JAN-2000
DEFINITION  xv20g11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2813732 3', mRNA sequence.
ACCESSION   AW303579
VERSION     AW303579.1  GI:6713268
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2813732"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      203 a    122 c     92 g    181 t
ORIGIN
```

```
  Query Match             5.9%;  Score 594.8;  DB 113;  Length 598;
  Best Local Similarity   99.7%;  Pred. No. 3.6e-118;
  Matches  596;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    7974 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 8033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 AGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGA 539

Qy    8034 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 8093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 TAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGC 479

Qy    8094 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 8153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 ATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTC 419

Qy    8154 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 8213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 TTTGTTGAACCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATT 359

Qy    8214 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgcctg 8273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 ATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTG 299

Qy    8274 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaatattatt 8333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 AGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATT 239

Qy    8334 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 8393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 AGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACT 179

Qy    8394 gctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccag 8453
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 GCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAG 119

Qy    8454 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 8513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     118 TGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTT 59

Qy    8514 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggca 8571
           |||||||||||||||||||| |||||||||||||||||||||||||||| ||||
Db      58 GCCTTAAGTTGAGAGGAGGTCACCTTTAGCTACTGTCTTTGGTTTGAGAGCCAAGGCA 1


RESULT  10
BG115118
LOCUS       BG115118       964 bp    mRNA            EST       30-JAN-2001
DEFINITION  602315968F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416424 5',
            mRNA sequence.
ACCESSION   BG115118
VERSION     BG115118.1  GI:12608624
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium '(LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10147  row: g  column: 17
            High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4416424"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      327 a    201 c    274 g    162 t
ORIGIN


  Query Match             5.9%;  Score 592.2;  DB 173;  Length 964;
  Best Local Similarity   84.4%;  Pred. No. 1.4e-117;
  Matches  737;  Conservative    0;  Mismatches  128;  Indels    8;  Gaps    6;

Qy    2713 caaatgccgtcagtgcagttttacagctgccgatactcagtcactactggagcacttcaa 2772
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAAATGCCGTCAGTGCAGTTTTACAGCTGCCGATACTCAGTCACTACTGGAGCACTTCAA 60

Qy    2773 cactgttcactgccaggaacaggacatcactacagccaacggcgaagaggacggtcatgc 2832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CACTGTTCACTGCCAGGAACAGGACATCACTACAGCCAACGGCGAAGAGGACGGTCATGC 120

Qy    2833 catatccaccatcaaagaggagcccaaaattgacttcagggtctacaatctgctaactcc 2892
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CATATCCACCATCAAAGAGGAGCCCAAAATTGACTTCAGGGTCTACAATCTGCTAACTCC 180

Qy    2893 agactctaaaatgggagagccagtttctgagagtgtggtgaagagagagaagctggaaga 2952
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGACTCTAAAATGGGAGAGCCAGTTTCTGAGAGTGTGGTGAAGAGAGAGAAGCTGGAAGA 240

Qy    2953 gaaggacgggctcaaagagaaagtttggaccgagagttccagtgatgaccttcgcaatgt 3012
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAAGGACGGGCTCAAAGAGAAAGTTTGGACCGAGAGTTCCAGTGATGACCTTCGCAATGT 300

Qy    3013 gacttggagaggggcagacatcctgcgggggagtccgtcatacacccaagcaagcctggg 3072
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GACTTGGAGAGGGGCAGACATCCTGCGGGGGAGTCCGTCATACACCCAAGCAAGCCTGGG 360

Qy    3073 gctgctgacgcctgtgtctggcacccaagagcagacaaagactctaagggatagtcccaa 3132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCTGCTGACGCCTGTGTCTGGCACCCAAGAGCAGACAAAGACTCTAAGGGATAGTCCCAA 420

Qy    3133 tgtggaggccgcccatctggcgcgacctatttatggcttggctgtgga-aaccaagggat 3191
           |||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||
Db     421 TGTGGAGGCCG-CCATCTGGCGCGACCTATTTATGGCTTGGCTGTGGACAACCAAGGGAT 479

Qy    3192 tcctgcaggggcgccagctggcggagagaagtctggggccctccccca-gcagtatcct 3250
           |||||| |||||||||||||||||||||||||||||||||||||| |||||| || ||||||||||
Db     480 TCCTGCCGGGGGCGCCAGCTGGCGGAGAGAAGTCTGGGGACCTCCCACAGGCAGTATCCT 539

Qy    3251 gcatcgggagaaaacaagtccaaggatgaatcccagtccctgttacggaggcgtagaggc 3310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 GCATCGGGAGAAAACAAGTCCAAGGATGAATCCCAGTCCCTGTTACGGAGGCGTAGAGGC 599

Qy    3311 tccggtgtttttttgtgccaattgcctgaccacaaagacctctctctggcgaaagaatgca 3370
           ||||  | |  | |||||||||| |||||||||||| |||||||||||  |   |||||
Db     600 TCCG--GGTAATAGTGCCAATTGGCTGACCACAAAGA-CTCTCTCTGGCCGACAGATGCA 656

Qy    3371 aatggcggatatgtatgcaacgcgtgtggcctctaccagaagcttcactcgactcccagg 3430
           |  |  |          || || ||| | || | ||||||||||  |   |||  |  |
Db     657 ATGGCGGATATGTATGACACCGAGTGGGCCCACAACCAGAAGCTCAA--TGACCACAGGC 714

Qy    3431 cctttaaacatcattaaacaaaacaacggtgagcagattattaggaggagaacaagaaag 3490
           |    | |  ||||||   || | |  ||||     | |  | ||  || | ||
Db     715 TTAAAACTATTTAACAAACAAGGTAAAGATATACAGAGGGAAAAGGAGGGAGTAACAGAG 774
```

```
Qy   3491 cgccttaacccagaggcacttcaggctgagcagctcaacaaacagcagaggggcagcaat 3550
          |          ||  | |  |||||  ||  ||   |  ||  | | || ||  |  |
Db    775 CATAGGGGGAAAGCAGAAAGTCAGGGAGAAGAGAAAACTAAGAAACGGAAGGAGAAGAGA 834

Qy   3551 gaggagcaagtcaatggaagcccgttagagagg 3583
          |  ||  |    || | ||    | | | | |
Db    835 TATAAGTTACGAGATAGTAGAGAATAATATACG 867


RESULT  11
BG178176
LOCUS       BG178176      917 bp    mRNA            EST       06-FEB-2001
DEFINITION  602327784F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429125 5',
            mRNA sequence.
ACCESSION   BG178176
VERSION     BG178176.1  GI:12684879
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 917)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10180  row: h  column: 22
            High quality sequence stop: 635.
FEATURES             Location/Qualifiers
     source          1. .917
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429125"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      287 a    186 c    172 g    272 t
ORIGIN


  Query Match             5.9%;  Score 587.6;  DB 174;  Length 917;
  Best Local Similarity   92.4%;  Pred. No. 1.4e-116;
  Matches  719;  Conservative    0;  Mismatches   44;  Indels   15;  Gaps    9;

Qy   5409 acaaatgtcagaccgtgggtttaatttgtttagaacacatggcatttcttcacaaggtaa 5468
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACAAATGTCAGACCG-GGGTTTAATTTGTTTAGAACACATGGCATTTCTTCACAAGGTAA 59

Qy   5469 cctgctgtatttatttatttttcttttggttaaatataatttccaaacttttgtggtcaggc 5528
          |||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db     60 CCTGCTGTATTTATTTATTTTC-TTTGGTTAAATATAATTTCCAAAC-TTGTGGTCAGGC 117

Qy   5529 agcgtctaaggttacgttaccacagactgacagttggtatatgtaccagccaatcccttc 5588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 AGCGTCTAAGGTTACGTTACCACAGACTGACAGTTGGTATATGTACCAGCCAATCCCTTC 177

Qy   5589 attaaatgtatacagatttagttaagtagcattaaataggattcttagaagtatgtcctc 5648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 ATTAAATGTATACAGATTTAGTTAAGTAGCATTAAATAGGATTCTTAGAAGTATGTCCTC 237

Qy   5649 atagaacttttaatacttaaggctttgtaaaaactatccatgaagggaaagctcctcagc 5708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 ATAGAACTTTTAATACTTAAGGCTTTGTAAAAACTATCCATGAAGGGAAAGCTCCTCAGC 297

Qy   5709 ataactgctcagggaaatagggctaaataactgaacattaaataattggttaaaggtgct 5768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 ATAACTGCTCAGGGAAATAGGGCTAAATAACTGAACATTAAATAATTGGTTAAAGGTGCT 357

Qy   5769 gttagtcgagcctcaatgcttgctacaaggatgtatgtacaaggactgactttaataatt 5828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 GTTAGTCGAGCCTCAATGCTTGCTACAAGGATGTATGTACAAGGACTGACTTTAATAATT 417

Qy   5829 tgcattatattgtcccaaccagtagtttatttttgccacggagatgtagaagatattac 5888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 TGCATTATATTGTCCCAACCAGTAGTTTATTTTTTGCCACGGAGATGTAGAAGATATTAC 477

Qy   5889 aagctactggatgcactgtcagattaacttatttcattaaagaagttgggagaacaaata 5948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 AAGCTACTGGATGCACTGTCAGATTAACTTATTTCATTAAAGAAGTTGGGAGAACAAATA 537

Qy   5949 ggaaaaaaaaaacttatttttctagtaaatattaatgtattacatttc-aaataatggtgc 6007
          |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db    538 GGAAAAAAAAAATTATTTTTCTAGTAAATATTAATGTATTACATTTCAAAATAATGGTGC 597

Qy   6008 ctgacatattgaataattattt-tctacagtgtacgtat-----gcaacaaagatattcc 6061
          |||||||||||||||||||| ||||    ||||||||      ||| | | ||   |
Db    598 CTGACATATTGAATAATTATTTCTCTAACAGGTACGTATTGCCAACAAAAGATATTCCCA 657

Qy   6062 atcatgcatt---agagtcagttctggctctgcctagctg-tttacatttgcaaatgtag 6117
          |||||||||    |   |  | |||| || ||||| ||| ||||||| ||  ||||| |
Db    658 TTCATGCATTTAGAAGTCCGCTCCTGGGTCGGCCTAACTGTTTTACATCTGGCAATGTCG 717

Qy   6118 caaacaaggtaatgaagcaactatttctattgcagtagatatccttttgtgtgtgtgt 6175
          || || |||  |||||   |||| |||||| |||| | | ||||||| | | | | ||
Db    718 CACACCAGGGTATGAACGCACTA-TTCTATGGCAGGAAAAATCCTTTCGGGGGCGGGT 774


RESULT  12
BF138096
LOCUS       BF138096      953 bp    mRNA            EST       24-OCT-2000
DEFINITION  601785006F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4012945 5',
            mRNA sequence.
ACCESSION   BF138096
VERSION     BF138096.1  GI:10977136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 953)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9254  row: l  column: 02
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1. .953
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4012945"
```

/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; transgenic model WNT-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 308 a 138 c 204 g 303 t
ORIGIN

Query Match 5.9%; Score 587.2; DB 144; Length 953;
Best Local Similarity 89.0%; Pred. No. 1.7e-116;
Matches 680; Conservative 0; Mismatches 78; Indels 6; Gaps 4;

```
Qy   5495 ggttaaatataatttccaaactttgtggtcaggcagcgtctaaggttacgttaccacaga 5554
          |||||||||||||| |  |||| ||||||||||||||| |  ||||||||||||||||||||
Db      1 GGTTAAATATAATGTGGAAAC-TTGTGGTCAGGCAGCATGCAAGGTTACGTTACCACAGA 59

Qy   5555 ctgacagttggtatatgtaccagccaatccccttcattaaatgtatacagatttagttaag 5614
          ||||||| ||||||||||||||  |||||||||||||| || ||||||||||||||||||||
Db     60 CTGACAG-TGGTATATGTACCAATCAATCCCTTCATTAGATTTATACAGATTTAGTTAAG 118

Qy   5615 tagcattaaataggattcttagaagtatgtcctcatagaacttttaatacttaaggcttt 5674
          ||||||||| |||| |||||||||||||||| ||||||  ||||||||||||||||||||
Db    119 TAGCATTAAGTAGGCTTCTTAGAAGTATGTCTTCATAGTCCTTTTAATACTTAAGGCTTT 178

Qy   5675 gtaaaaactatccatgaagggaaagctcctcagcataactgctcagggaaatagggctaa 5734
          |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db    179 GTAAAAACTATCCATGAAGGGAGAGCTCCTCGGCATAACTGCTCAGGGAAATAGGGCTAA 238

Qy   5735 ataactgaacattaaataattggttaaaggtgctgttagtcgagcctcaatgcttgctac 5794
          |||| |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    239 ATAATTGAACATTAAATAATTGGTTTAAGGTGCTGTTAGTCGAGCCTCAATGCTTGCTAC 298

Qy   5795 aaggatgtatgtacaaggactgactttaataatttgcattatattgtcccaaccagtagt 5854
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 AAGGATGTATGTACAAGGACTGACTTTAATAATTTGCATTATATTGTCCCAACCAGTAGT 358

Qy   5855 ttatttttgccacggagatgtagaagatattacaagctactggatgcactgtcagatta 5914
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 TTATTTTTTGCCACGGAGATGTAGAAGATATTACAAGCTACTGGATGCACTGTCAGATTA 418

Qy   5915 acttatttcattaaagaagttgggagaacaaataggaaaaaaaaaacttatttttctagta 5974
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 ACTTATTTCATTAAAGAAGTTGGGAGAACAAATAGGAAAAAAAAACTTATTTTTCTAGTA 478

Qy   5975 aatattaatgtattacatttcaaataatggtgcctgacatat-tgaataattatttttcta 6033
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    479 AATATTAATGTATTACATTTCAAATAATGGTGCCTGACATATGTGAATAATTATTTTTCTA 538

Qy   6034 cagtgtacgtatgcaacaaagatattccatcatgcattagagtcagttctggctctgcct 6093
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 CAGTGTACGTATGCAACAAAGATATTCCATCATGCATTAGAGTCAGTTCTGGCTCTGCCT 598

Qy   6094 agctgtttacatttgcaaatgtagcaaacaaggtaatgaagcaactatttctattgcagt 6153
          ||||||||||||||||||||||||||||||||||||||||    | | |   | ||
Db    599 AGCTGTTTACATTTGCAAATGTAGCAAACAAGGTAATGAAGCACTATTCTATGCAGTAGG 658

Qy   6154 agatatccttttgtgtgtgtg---tgtgtgcattaaagttgtaaacggtaacatgaaaca 6210
               | ||| ||||||   |||||| ||    |   |||   ||||  | | |
Db    659 TATCTCTAGTGTGTATGTGTGAGCTGTGTGTATGTGTGCATTAACGTGTAAGGTAACATG 718

Qy   6211 aatgaaagttcttgctataatggtatggaaaacaagaaggaaat 6254
          ||  || |   |  |    ||||| |||| |   || | || ||
Db    719 AAACAAGGAGTTCTCAGGTATGGTGTGGACACAGAGCATGACAT 762
```

RESULT 13
AG025673/c
LOCUS AG025673 602 bp DNA GSS 21-DEC-1999
DEFINITION Homo sapiens DNA, chromosome 8q23, reverse end of BAC clone:
KB1432F7, genomic survey sequence.
ACCESSION AG025673
VERSION AG025673.1 GI:6624364
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Homo sapiens DNA, clone_lib:KEIO BAC library clone:KB1432F7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS Shimizu,N. and Asakawa,S.
TITLE The BAC end sequence of Homo sapiens BAC clone KB1432F7
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 602)
AUTHORS Shimizu,N. and Asakawa,S.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) to the DDBJ/EMBL/GenBank databases. Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES Location/Qualifiers
source 1. .602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="KB1432F7"
/clone_lib="KEIO BAC library"
/map="8q23"
/note="This sequence is reverse end of BAC clone KB1432F7."
BASE COUNT 190 a 127 c 87 g 195 t 3 others
ORIGIN

Query Match 5.7%; Score 571.8; DB 229; Length 602;
Best Local Similarity 98.8%; Pred. No. 3.5e-113;
Matches 595; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```
Qy   8031 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaca 8090
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    602 AGATAGTNATGTTCTGTTCTAAACTGTTCTGTTTAAGTGAAGGTAGATTTTTATAAANCA 543

Qy   8091 agcatggggattcttttctaaggtaatattaatgagaaggg-aaaaaagtatctttaaca 8149
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db    542 AGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAAGTATCTTTANCA 483

Qy   8150 gctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatc 8209
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GCTCTTTGTTAAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATC 423

Qy   8210 tattatgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatg 8269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 TATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATG 363

Qy   8270 cctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataat 8329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CCTGAGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAAT 303

Qy   8330 tattagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtg 8389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TATTAGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTG 243

Qy   8390 aactgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccat 8449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AACTGCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCAT 183

Qy   8450 ccagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcc 8509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCC 123

Qy   8510 ttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatgg 8569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGG 63
```

```
Qy   8570 c-aaaaaaaaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagc 8628
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CAAAAAAAAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGC 3

Qy   8629 tt 8630
          ||
Db      2 TT 1


RESULT  14
BF139707
LOCUS       BF139707      749 bp    mRNA            EST       24-OCT-2000
DEFINITION  601785355F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4013365 5',
            mRNA sequence.
ACCESSION   BF139707
VERSION     BF139707.1  GI:10978747
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9255  row: m  column: 14
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4013365"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      210 a    181 c    187 g    171 t
ORIGIN


  Query Match             5.7%;  Score 567.8;  DB 144;  Length 749;
  Best Local Similarity   86.1%;  Pred. No. 2.6e-112;
  Matches  640;  Conservative    0;  Mismatches  102;  Indels    1;  Gaps    1;

Qy   2118 caatgaccaagacagacaagagctcgagtggggcta-aaaagaaggacttctccagcaag 2176
          || |||||||| |||| ||| ||| |||||||||||| || |||||||||||| ||||||||
Db      2 CATTGACCAAGCCAGAAAAGGGCTTGAGTGGGGCTATAAGAGAAGGACTTCCCCAGCAAG 61

Qy   2177 ggagccgaggataatatggtaacgagctataattgtcagttctgtgacttccgatattcc 2236
          ||||| |||||||||||||||||||||||||| ||||||||||||||||  ||||||||
Db     62 GGAGCAGAGGATAATATGGTAACGAGCTATAACTGTCAGTTCTGTGACTTTAGATATTCC 121

Qy   2237 aaaagccatggccctgatgtaattgtagtggggccacttctccgtcattatcaacagctc 2296
          |||||||||||||||||| |||||||||| |||||||||||||||||| || |||
Db    122 AAAAGCCATGGCCCTGATGTGATTGTAGTGGGTCCACTTCTCCGTCATTATCAGCAACTC 181

Qy   2297 cataacattcacaagtgtaccattaaacactgtccattctgtcccagaggactttgcagc 2356
          |||||||| |||||||||||||||| |||||||||||||||||||||||| |||||||||
Db    182 CATAACATCCACAAGTGTACCATTAAGCACTGTCCATTCTGTCCCAGAGGCCTTTGCAGC 241

Qy   2357 ccagaaaagcaccttggagaaattacttatccgtttgcttgtagaaaaagtaattgttcc 2416
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    242 CCAGAAAAGCACCTTGGAGAGATTACTTATCCGTTTGCTTGTAGAAAAAGTAATTGTTCC 301

Qy   2417 cactgtgcactcttgcttctgcacttgtctcctggggcggctggaagctcgcgagtcaaa 2476
          |||||||||| |||||  |||||||||||||||||| ||| ||||| ||| ||||||||
Db    302 CACTGTGCACTGTTGCTGTTGCACTTGTCTCCTGGGGTGGCCGGAAGTTCGAGAGTCAAA 361

Qy   2477 catcagtgccatcagtgttcattcaccacccctgacgtagatgtactcctctttcactat 2536
          || || ||||| |||||||| ||| | ||||| || |||||||| ||||||||||||||||
Db    362 CACCAATGCCACCAGTGTTCCTTCTCTACCCCCGATGTAGATGTGCTCCTCTTTCACTAT 421

Qy   2537 gaaagtgtgcatgagtcccaagcatcggatgtcaaacaagaagcaaatcacctgcaagga 2596
          || | ||||||||||||||||||||||||||||||| || || || ||||||| ||||
Db    422 GAGACCGTGCATGAGTCCCAAGCATCGGATGTCAAACAGGAGGCCAACCACCTGCTAGGA 481

Qy   2597 tcggatgggcagcagtctgtcaaggaaagcaaagaacactcatgtaccaaatgtgatttt 2656
          || || ||||||||| ||| || ||| ||||| || ||||| || |||||||| ||||||
Db    482 TCTGACGGGCAGCAGGCTGCCAGGGACAGCAAGGAGCACTCGTGCACCAAATGCGATTTT 541

Qy   2657 attacccaagtggaagaagagatttcccgacactacaggagagcacacagctgctacaaa 2716
          || ||||| ||||| |||||||| ||  ||||||||||||||| |||||||||| | |
Db    542 ATAACCCAGGTGGACGAAGAGATCTCGAGACACTACAGGAGAGCAGACAGCTGCTATACA 601

Qy   2717 tgccgtcagtgcagttttacagctgccgatactcagtcactactggagcacttcaacact 2776
          || |||||||||||||| |||||||| || | |||||||||  |||||||||||| ||
Db    602 TGTCGTCAGTGCAGTTTCACAGCTGCTGACAGTCAGTCACTATAGGAGCACTTCAAGACA 661

Qy   2777 gttcactgccaggaacaggacatcactacagccaacggcgaagaggacggtcatgccata 2836
          |||||  | |||||||||||| |||| ||| |  || || ||||| |   | |    |||
Db    662 GTTCATGGGCAGGAACAGGAAATCATTACTGGTAATGGGGAAGAAGGGTGGCCATGCATT 721

Qy   2837 tccaccatcaaagaggagcccaa 2859
          ||| ||| ||  ||   |  ||
Db    722 ACCAACATTAAGAAGAGCCAAAA 744


RESULT  15
AW021592
LOCUS       AW021592      569 bp    mRNA            EST       13-SEP-1999
DEFINITION  df25f07.y1 Morton Fetal Cochlea Homo sapiens cDNA clone
            IMAGE:2484612 5', mRNA sequence.
ACCESSION   AW021592
VERSION     AW021592.1  GI:5875122
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
            and Morton,C.C.
  TITLE     Isolation of novel and known genes from a human fetal cochlear cDNA
            library using subtractive hybridization and differential screening
  JOURNAL   Genomics 23, 42-50 (1994)
  MEDLINE   95130111
COMMENT     Contact: Morton, C. C.
            Departments of Pathology and Obstetrics, Gynecology and
            Reproductive Biology
            Brigham and Women's Hospital
            75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
            Tel: 617 732 7980
            Fax: 617 738 6996
            Email: ccmorton@bics.bwh.harvard.edu
            DNA sequencing and analyses were performed by National Institutes
            of Health Intramural Sequencing Center (NISC; see
            http://www.nisc.nih.gov).
             This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Plate: LLAM6182  row: K  column: 13
```

```
           Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2484612"
                     /clone_lib="Morton Fetal Cochlea"
                     /tissue_type="cochlea"
                     /dev_stage="16-22 week fetus"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI;
                     Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
                     unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
                     37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
                     XR Vector. Library constructed by N. Robertson, C. Morton.
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      182 a     95 c     94 g    198 t
ORIGIN


  Query Match             5.6%;  Score 557;  DB 32;  Length 569;
  Best Local Similarity   99.8%;  Pred. No. 5.5e-110;
  Matches  568;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy    7722 attatttgtctatattctaaattgagctttggatcaaactttaggccaggaccagctcat 7781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATTATTTGTCTATATTCTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCAT 60

Qy    7782 gcgttctcattcttcctttctcactctttctctcatcactcacctctgtattcattctg 7841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCGTTCTCATTCTTCCTTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTG 120

Qy    7842 ttgtttgggatagaaaaatcataaagagccaacccatctcagaacgttgtggattgagag 7901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTGTTTGGGATAGAAAAATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAG 180

Qy    7902 agacactacatgactccaagtatatgagaaaaggacagagctctaattgataactctgta 7961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGACACTACATGACTCCAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTA 240

Qy    7962 gttcaaaagg-aaagagtatgcccaattctctctacatgacatattgagatttttttaa 8020
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTTCAAAAGGAAAAGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAA 300

Qy    8021 tcaacttttaagatagtgatgttctgttctaaactgttctgttttagtgaaggtagattt 8080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCAACTTTTAAGATAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTT 360

Qy    8081 ttataaaacaagcatggggattcttttctaaggtaatattaatgagaagggaaaaaagta 8140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTATAAAACAAGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTA 420

Qy    8141 tctttaacagctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtg 8200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCTTTAACAGCTCTTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTG 480

Qy    8201 cacataatctattatgatccaatgcaaatacagctccaaaaatattaaatgtatatatat 8260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CACATAATCTATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATAT 540

Qy    8261 tttaaaatgcctgaggaaatacatttttc 8289
           |||||||||||||||||||||||||||||
Db     541 TTTAAAATGCCTGAGGAAATACATTTTTC 569


Search completed: April 30, 2001, 12:18:55
Job time: 12604 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 18:55:22 ; Search time 22667.1 Seconds
(without alignments)
632.171 Million cell updates/sec

Title: US-09-721-183-3
Perfect score: 2327
Sequence: 1 gatggatgcatctcaaaatg.....actcaattctggttttcccc 2327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2257.6	97.0	10011	88	AF183810	AF183810 Homo sapi
2		2253	96.8	167123	63	AC013815	AC013815 Homo sapi
3	c	2237.8	96.2	159505	68	AC023142	AC023142 Homo sapi
4		1584.4	68.1	1587	89	AK021511	AK021511 Homo sapi
5		240	10.3	674	10	AX067327	AX067327 Sequence
6	c	86.2	3.7	224448	84	PFMAL4P4	AL035477 Plasmodiu
7		82.6	3.5	879	53	CNS01JRG	AL147405 Anopheles
8		80.2	3.4	53932	68	AC023371	AC023371 Homo sapi
9	c	79.2	3.4	936	54	CNS07ANM	AL436808 T7 end of
10		78.6	3.4	1084	53	CNS06PLB	AL409509 T7 end of
11		76.8	3.3	1169	53	CNS06KHQ	AL402900 T3 end of

	12	76.4	3.3	897	54	CNS07ABZ
	13	76.2	3.3	60604	68	AC023466
c	14	76.2	3.3	130540	76	AC079417
	15	75.8	3.3	83440	69	AC024285
c	16	75.4	3.2	192929	60	AC005505
c	17	75.4	3.2	256172	60	AC005139
c	18	75.4	3.2	310779	60	AC005140
	19	75	3.2	187279	70	AC026698
c	20	74.6	3.2	22243	6	PFAVAR23A
c	21	74.6	3.2	127709	83	CNS05TC1
	22	74.4	3.2	115218	91	HS159A1
c	23	73.6	3.2	759	53	CNS06QXV
	24	73.6	3.2	99003	81	AL390756
	25	73.2	3.1	12600	63	AC013952
	26	73.2	3.1	88932	67	AC022680
	27	73.2	3.1	156550	64	AC015830
	28	73.2	3.1	256172	60	AC005139
	29	73	3.1	840	14	CNS018QK
c	30	73	3.1	152209	91	HS1108D11
c	31	72.6	3.1	1032	53	CNS06LOP
c	32	72.6	3.1	108908	6	PFMAL3P8
c	33	72.4	3.1	61864	6	CEY50E8A
c	34	72.4	3.1	195969	76	AC074354
	35	72.2	3.1	81120	68	AC022851
c	36	71.8	3.1	161891	60	AC008206
c	37	71.6	3.1	192465	83	AP002851
	38	71.4	3.1	51014	62	AC011856
	39	71.2	3.1	70914	78	AC087457
c	40	71.2	3.1	153936	77	AC079863
c	41	71.2	3.1	173915	90	AP001254
c	42	71.2	3.1	340000	90	AP001678
c	43	71	3.1	104992	60	AC005504
c	44	71	3.1	167123	63	AC013815
c	45	71	3.1	169546	60	AC004157

AL436389	T7 end of
AC023466	Homo sapi
AC079417	Mus muscu
AC024285	Homo sapi
AC005505	Plasmodiu
AC005139	Plasmodiu
AC005140	Plasmodiu
AC026698	Homo sapi
L40609	Plasmodium
AL352983	Homo sapi
AL034397	Human DNA
AL411257	T7 end of
AL390756	Homo sapi
AC013952	Drosophil
AC022680	Homo sapi
AC015830	Homo sapi
AC005139	Plasmodiu
AL110675	Botrytis
AL034419	Human DNA
AL404447	T7 end of
AL034560	Plasmodiu
AL117200	Caenorhab
AC074354	Oryza sat
AC022851	Homo sapi
AC008206	Drosophil
AP002851	Homo sapi
AC011856	Homo sapi
AC087457	Homo sapi
AC079863	Homo sapi
AP001254	Homo sapi
AP001678	Homo sapi
AC005504	Plasmodiu
AC013815	Homo sapi
AC004157	Plasmodiu

ALIGNMENTS

```
RESULT   1
AF183810
LOCUS       AF183810    10011 bp    mRNA            PRI       10-JAN-2000
DEFINITION  Homo sapiens zinc finger transcription factor TRPS1 mRNA, complete
            cds.
ACCESSION   AF183810
VERSION     AF183810.1  GI:6684533
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10011)
  AUTHORS   Momeni,P., Glockner,G., Schmidt,O., von Holtum,D., Albrecht,B.,
            Gillessen-Kaesbach,G., Hennekam,R., Meinecke,P., Zabel,B.,
            Rosenthal,A., Horsthemke,B. and Ludecke,H.J.
  TITLE     Mutations in a new gene, encoding a zinc-finger protein, cause
            tricho-rhino-phalangeal syndrome type I
  JOURNAL   Nat. Genet. 24 (1), 71-74 (2000)
  MEDLINE   20082812
REFERENCE   2  (bases 1 to 10011)
  AUTHORS   Momeni,P., Gloeckner,G., Schmidt,O., von Holtum,D., Albrecht,B.,
            Gillessen-Kaesbach,G., Hennekam,R., Meinecke,P., Zabel,B.,
            Rosenthal,A., Horsthemke,B. and Luedecke,H.-J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1999) Institut fuer Humangenetik,
            Universitaetsklinikum, Hufelandstr. 55, D-45122 Essen, Germany
FEATURES             Location/Qualifiers
     source          1. .10011
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q24.1"
     CDS             639. .4484
                     /note="tricho-rhino-phalangeal syndrome type I protein"
                     /codon_start=1
                     /product="zinc finger transcription factor TRPS1"
                     /protein_id="AAF23614.1"
                     /db_xref="GI:6684534"
                     /translation="MVRKKNPPLRNVASEGEGQILEPIGTESKVSGKNKEFSADQMSE
                     NTDQSDAAELNHKEEHSLHVQDPSSSSKKDLKSAVLSEKAGFNYESPSKGGNFPSFPH
                     DEVTDRNMLAFSFPAAGGVCEPLKSPQRAEADDPQDMACTPSGDSLETKEDQKMSPKA
                     TEETGQAQSGQANCQGLSPVSVASKNPQVPSDGGVRLNKSKTDLLVNDNPDPAPLSPE
                     LQDFKCNICGYGYYGNDPTDLIKHFRKYHLGLHNRTRQDAELDSKILALHNMVQFSHS
                     KDFQKVNRSVFSGVLQDINSSRPVLLNGTYDVQVTSGGTFIGIGRKTPDCQGNTKYFR
                     CKFCNFTYMGNSSTELEQHFLQTHPNKIKASLPSSEVAKPSEKNSNKSIPALQSSDSG
                     DLGKWQDKITVKAGDDTPVGYSVPIKPLDSSRQNGTEATSYYWCKFCSFSCESSSSLK
                     LLEHYGKQHGAVQSGGLNPELNDKLSRGSVINQNDLAKSSEGETMTKTDKSSSGAKKK
                     DFSSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLHNIHKCTIKHCPF
                     CPRGLCSPEKHLGEITYPFACRKSNCSHCALLLLHLSPGAAGSSRVKHQCHQCSFTTP
                     DVDVLLFHYESVHESQASDVKQEANHLQGSDGQQSVKESKEHSCTKCDFITQVEEEIS
                     RHYRRAHSCYKCRQCSFTAADTQSLLEHFNTVHCQEQDITTANGEEDGHAISTIKEEP
                     KIDFRVYNLLTPDSKMGEPVSESVVKREKLEEKDGLKEKVWTESSSDDLRNVTWRGAD
                     ILRGSPSYTQASLGLLTPVSGTQEQTKTLRDSPNVEAAHLARPIYGLAVETKGFLQGA
                     PAGGEKSGALPQQYPASGENKSKDESQSLLRRRRGSGVFCANCLTTKTSLWRKNANGG
                     YVCNACGLYQKLHSTPRPLNIIKQNNGEQIIRRRTRKRLNPEALQAEQLNKQQRGSNE
                     EQVNGSPLERRSEDHLTESHQREIPLPSLSKYEAQGSLTKSHSAQQPVLVSQTLDIHK
                     RMQPLHIQIKSPQESTGDPGNSSSVSEGKGSSERGSPIEKYMRPAKHPNYSPPGSPIE
                     KYQYPLFGLPFVHNDFQSEADWLRFWSKYKLSVPGNPHYLSHVPGLPNPCQNYVPYPT
                     FNLPPHFSAVGSDNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVKTEK
                     VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHGDSGPFQCSICQHLCTDKYDFTTHI
                     QRGLHRNNAQVEKNGKPKE"
BASE COUNT     3067 a    1963 c    2056 g    2925 t
ORIGIN
```

```
  Query Match             97.0%;  Score 2257.6;  DB 88;  Length 10011;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 2293;  Conservative    0;  Mismatches   32;  Indels    3;  Gaps    3;

Qy       1 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6898 GATGGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAA 6957

Qy      61 agagaggagattcccccaaacaacaatatttaatttttcttagtaaaaagaataacagaat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6958 AGAGAGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAAT 7017

Qy     121 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7018 GCATCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACC 7077

Qy     181 agaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactcatt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7078 AGAAGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATT 7137

Qy     241 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7138 AAGCAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTT 7197

Qy     301 tttattttcattcaatttttttttttttctcagacagacatagtagtaacgactagcattg 360
          |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db    7198 TTTATTTTCATTCAATTTTTTTTTTTTCTCAGACAGACATAGTAGTATCAACTAGCATTG 7257

Qy     361 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtagaatatt 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7258 GAAAATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAAAATATT 7317

Qy     421 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7318 TTTGGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTAT 7377

Qy     481 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7378 TTTCATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTG 7437

Qy     541 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 600
```

```
Db   7438 AAAGAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTA 7497

Qy    601 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 660
Db   7498 TTCCATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATT 7557

Qy    661 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 720
Db   7558 TTATCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTAATGAGGAA 7617

Qy    721 tttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgat 780
Db   7618 TTTCATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGAT 7677

Qy    781 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 840
Db   7678 GTGCCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTAT 7737

Qy    841 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 900
Db   7738 ATTCTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCT 7797

Qy    901 tccttttctcactctttctctcancactcacctctgtattcattctgttgtttgggatag 960
Db   7798 TCCTTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAG 7857

Qy    961 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 1020
Db   7858 AAAAATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGA 7917

Qy   1021 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 1080
Db   7918 CTCCAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAA 7977

Qy   1081 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 1140
Db   7978 AGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGA 8037

Qy   1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
Db   8038 TAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGC 8097

Qy   1201 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
Db   8098 ATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTC 8157

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 1320
Db   8158 TTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATT 8217

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgcctg 1380
Db   8218 ATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTG 8277

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
Db   8278 AGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATT 8337

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
Db   8338 AGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACT 8397

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccag 1560
Db   8398 GCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAG 8457

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 1620
Db   8458 TGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTT 8517

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcnnn 1680
Db   8518 GCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAA 8577

Qy   1681 nnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagcttgct 1740
Db   8578 AAAAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGCT 8637

Qy   1741 gactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaa 1800
Db   8638 GACTGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTAA 8697

Qy   1801 atattctaatcaaaggtaggtactgttaggtggaatttctatcagcaggcaactgcaaatg 1860
Db   8698 ATATTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAATG 8757

Qy   1861 agaagaagatagaaggacgcccgtcgggactttggagggcagtgttattttcccaaagaa 1920
Db   8758 AGAAGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCATTGTTATTTTCCCAAAGAA 8817

Qy   1921 agacggccaagggcagaggcatggattctttgcagagcacttcctttttggtttttcagta 1980
Db   8818 AGACGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTA 8877

Qy   1981 ctgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcat 2040
Db   8878 CTGTTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCAT 8937

Qy   2041 cccagtt-attgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagta 2099
Db   8938 CCCAGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTA 8997

Qy   2100 tgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgt 2159
Db   8998 TGTCACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATGT 9057

Qy   2160 acattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacacc 2219
Db   9058 ACATTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACC 9117

Qy   2220 ggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatgcagagaaaat 2279
Db   9118 GGAGAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTAT-CAGAGAAAAT 9176

Qy   2280 tgtccttcctagagcatagagtcccaaaactcaattctggttttcccc 2327
Db   9177 TGTCCTT-CTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCC 9223
```

```
RESULT   2
AC013815
LOCUS       AC013815   167123 bp    DNA             HTG      07-JUN-2000
DEFINITION  Homo sapiens clone RP11-21J16, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC013815
VERSION     AC013815.4  GI:8316201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167123)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
```

```
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Jun 7, 2000 this sequence version replaced gi:7341965.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4033
                Center clone name: 21_J_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160651 bases at least Q40
                Consensus quality: 163054 bases at least Q30
                Consensus quality: 164289 bases at least Q20
                Insert size: 167000; agarose-fp
                Insert size: 166123; sum-of-contigs
                Quality coverage: 6.6 in Q20 bases; agarose-fp
                Quality coverage: 6.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1116: contig of 1116 bp in length
            *    1117 1216: gap of      100 bp
            *    1217     2561: contig of 1345 bp in length
            *    2562 2661: gap of      100 bp
            *    2662     3754: contig of 1093 bp in length
            *    3755 3854: gap of      100 bp
            *    3855     4514: contig of 660 bp in length
            *    4515 4614: gap of      100 bp
            *    4615    11771: contig of 7157 bp in length
            *   11772 11871: gap of      100 bp
            *   11872    22787: contig of 10916 bp in length
            *   22788 22887: gap of      100 bp
            *   22888    31860: contig of 8973 bp in length
            *   31861 31960: gap of      100 bp
            *   31961    53736: contig of 21776 bp in length
            *   53737 53836: gap of      100 bp
            *   53837    81814: contig of 27978 bp in length
            *   81815 81914: gap of      100 bp
            *   81915   123726: contig of 41812 bp in length
            *  123727 123826: gap of      100 bp
            *  123827   167123: contig of 43297 bp in length.
  FEATURES             Location/Qualifiers
       source          1. .167123
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="RP11-21J16"
                       /clone_lib="RPCI-11 Human Male BAC"
       misc_feature    1. .1116
                       /note="assembly_fragment"
       misc_feature    1217. .2561
                       /note="assembly_fragment"
       misc_feature    2662. .3754
                       /note="assembly_fragment"
       misc_feature    3855. .4514
                       /note="assembly_fragment
                       clone_end:T7
                       vector_side:right"
       misc_feature    4615. .11771
                       /note="assembly_fragment"
       misc_feature    11872. .22787
                       /note="assembly_fragment"
       misc_feature    22888. .31860
                       /note="assembly_fragment
                       clone_end:SP6
                       vector_side:left"
       misc_feature    31961. .53736
                       /note="assembly_fragment"
       misc_feature    53837. .81814
                       /note="assembly_fragment"
       misc_feature    81915. .123726
                       /note="assembly_fragment"
       misc_feature    123827. .167123
                       /note="assembly_fragment"
  BASE COUNT     51442 a  29991 c  31207 g  53480 t   1003 others
  ORIGIN


    Query Match             96.8%;  Score 2253;  DB 63;  Length 167123;
    Best Local Similarity   98.6%;  Pred. No. 0;
    Matches 2297;  Conservative    0;  Mismatches   28;  Indels    4;  Gaps    4;

Qy       1 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28915 GATGGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAA 28974

Qy      61 agagaggagattcccccaaacaacaatatttaatttttcttagtaaaaagaataacagaat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28975 AGAGAGGAGATTCGCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAAT 29034

Qy     121 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29035 GCATCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACC 29094

Qy     181 agaagtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactcatt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29095 AGAAGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATT 29154

Qy     241 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29155 AAGCAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTT 29214

Qy     301 tttattttcattcaatttttttttttttctcagacagacatagtagtaacgactagcattg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29215 TTTATTTTCATTCAATTTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTG 29274

Qy     361 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtagaatatt 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29275 GAAAATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTTAGTAGAATATT 29334

Qy     421 tttggatagcgttgacacgatagatcttattccatacttctttattattgataattttat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29335 TTTGGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTAT 29394

Qy     481 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29395 TTTCATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTG 29454

Qy     541 aaagagacaaaaatttattataacactaaacactcctttttttgacatattaaagccttta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29455 AAAGAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTTGACATATTAAAGCCTTTA 29514

Qy     601 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29515 TTCCATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATT 29574

Qy     661 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 720
```

```
Db  29575 TTATCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTAATGAGGAA 29634

Qy    721 tttcatgatgatttaggaatttttctctcttggaaaaggcttcccctgtgatgaaaatgat 780
Db  29635 TTTCATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGAT 29694

Qy    781 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 840
Db  29695 GTGCCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTAT 29754

Qy    841 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 900
Db  29755 ATTCTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCT 29814

Qy    901 tccttttctcactctttctctcancactcacctctgtattcattctgttgtttgggatag 960
Db  29815 TCCTTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAG 29874

Qy    961 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 1020
Db  29875 AAAAATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGA 29934

Qy   1021 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 1080
Db  29935 CTCCAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAA 29994

Qy   1081 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 1140
Db  29995 AGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGA 30054

Qy   1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
Db  30055 TAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGC 30114

Qy   1201 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
Db  30115 ATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTC 30174

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 1320
Db  30175 TTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATT 30234

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctg 1380
Db  30235 ATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTG 30294

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
Db  30295 AGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATT 30354

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
Db  30355 AGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACT 30414

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccag 1560
Db  30415 GCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAG 30474

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 1620
Db  30475 TGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTT 30534

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc-nn 1679
Db  30535 GCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAA 30594

Qy   1680 nnnnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagcttgc 1739
Db  30595 AAAAAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGC 30654

Qy   1740 tgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaacta 1799
Db  30655 TGACTGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTA 30714

Qy   1800 aatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaat 1859
Db  30715 AATATTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAAT 30774

Qy   1860 gagaagaagatagaaggacgcccgtcgggactttggagggcagtgttatttttcccaaaga 1919
Db  30775 GAGAAGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCAGTGTTATTTTCCCAAAGA 30834

Qy   1920 aagacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagt 1979
Db  30835 AAGACGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGT 30894

Qy   1980 actgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagca 2039
Db  30895 ACTGTTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCA 30954

Qy   2040 tcccagtt-attgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagt 2098
Db  30955 TCCCAGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGT 31014

Qy   2099 atgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatg 2158
Db  31015 ATGTCACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATG 31074

Qy   2159 tacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacac 2218
Db  31075 TACATTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACAC 31134

Qy   2219 cggagaaaagcaggaataactttaccgtggaaataatgtttagctttttatgcagagaaaa 2278
Db  31135 CGGAGAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTAT-CAGAGAAAA 31193

Qy   2279 ttgtccttcctagagcatagagtcccaaaactcaattctggttttcccc 2327
Db  31194 TTGTCCTT-CTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCC 31241
```

```
RESULT   3
AC023142/c
LOCUS       AC023142   159505 bp    DNA              HTG       20-AUG-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-558B2, WORKING DRAFT SEQUENCE,
            19 unordered pieces.
ACCESSION   AC023142
VERSION     AC023142.4  GI:9795980
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159505)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159505)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 11, 2000 this sequence version replaced gi:9558667.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0558B02
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
```

```
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 152349 bases at least Q40
Consensus quality: 154057 bases at least Q30
Consensus quality: 154743 bases at least Q20
Insert size: 141000; agarose-fp
Insert size: 157705; sum-of-contigs
Quality coverage: 4.80 in Q20 bases; agarose-fp
Quality coverage: 4.85 in Q20 bases; sum-of-contigs
-----------------------------------------------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1432: contig of 1432 bp in length
*     1433     1532: gap of unknown length
*     1533     2904: contig of 1372 bp in length
*     2905     3004: gap of unknown length
*     3005     5152: contig of 2148 bp in length
*     5153     5252: gap of unknown length
*     5253     7331: contig of 2079 bp in length
*     7332     7431: gap of unknown length
*     7432    11987: contig of 4556 bp in length
*    11988    12087: gap of unknown length
*    12088    16583: contig of 4496 bp in length
*    16584    16683: gap of unknown length
*    16684    21578: contig of 4895 bp in length
*    21579    21678: gap of unknown length
*    21679    26807: contig of 5129 bp in length
*    26808    26907: gap of unknown length
*    26908    31857: contig of 4950 bp in length
*    31858    31957: gap of unknown length
*    31958    37710: contig of 5753 bp in length
*    37711    37810: gap of unknown length
*    37811    43884: contig of 6074 bp in length
*    43885    43984: gap of unknown length
*    43985    53419: contig of 9435 bp in length
*    53420    53519: gap of unknown length
*    53520    63008: contig of 9489 bp in length
*    63009    63108: gap of unknown length
*    63109    75765: contig of 12657 bp in length
*    75766    75865: gap of unknown length
*    75866    90579: contig of 14714 bp in length
*    90580    90679: gap of unknown length
*    90680   104807: contig of 14128 bp in length
*   104808   104907: gap of unknown length
*   104908   118444: contig of 13537 bp in length
*   118445   118544: gap of unknown length
*   118545   135079: contig of 16535 bp in length
*   135080   135179: gap of unknown length
*   135180   159505: contig of 24326 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159505
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-558B2"
     misc_feature    1. .1432
                     /note="assembly_name:Contig10"
     misc_feature    1533. .2904
                     /note="assembly_name:Contig11"
     misc_feature    3005. .5152
                     /note="assembly_name:Contig12"
     misc_feature    5253. .7331
                     /note="assembly_name:Contig13"
     misc_feature    7432. .11987
                     /note="assembly_name:Contig14"
     misc_feature    12088. .16583
                     /note="assembly_name:Contig15"
     misc_feature    16684. .21578
                     /note="assembly_name:Contig16"
     misc_feature    21679. .26807
                     /note="assembly_name:Contig17"
     misc_feature    26908. .31857
                     /note="assembly_name:Contig18"
     misc_feature    31958. .37710
                     /note="assembly_name:Contig19"
     misc_feature    37811. .43884
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:right"
     misc_feature    43985. .53419
                     /note="assembly_name:Contig21"
     misc_feature    53520. .63008
                     /note="assembly_name:Contig22"
     misc_feature    63109. .75765
                     /note="assembly_name:Contig23"
     misc_feature    75866. .90579
                     /note="assembly_name:Contig24"
     misc_feature    90680. .104807
                     /note="assembly_name:Contig25"
     misc_feature    104908. .118444
                     /note="assembly_name:Contig26"
     misc_feature    118545. .135079
                     /note="assembly_name:Contig27"
     misc_feature    135180. .159505
                     /note="assembly_name:Contig28"
BASE COUNT    50304 a  28279 c  27746 g  51369 t   1807 others
ORIGIN


  Query Match             96.2%;  Score 2237.8;  DB 68;  Length 159505;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 2294;  Conservative    0;  Mismatches   30;  Indels    5;  Gaps    5;

Qy       1 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66624 GATGGATGCATCTCAAAATGTATAGCCAGACTTGAGAGGTGACAATTAAAGATCTAAAAA 66565

Qy      61 agagaggagattcccccaaacaacaatatttaatttttcttagtaaaaagaataacagaat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66564 AGAGAGGAGATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAAT 66505

Qy     121 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66504 GCATCGTGGCAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACC 66445

Qy     181 agaagtttggttaacttgggcaatatgacaagtattactttttgggcaaaactactcatt 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66444 AGAAGTTTGGTTAACTTGGGCAATATGACAAGTATTACTTTTTGGGCAAAACTACTCATT 66385

Qy     241 aagcaatttctctagtgtgtcggacacaaataggttctttatttttgtgcatgtatgcctt 300
           ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db   66384 AAGCAATTTCTCTAGTGTGTCGGACACAAATAGGTTCTTTATTTTGTGCATGTATGCCTT 66325

Qy     301 tttattttcattcaatttttttttttttctcagacagacatagtagtaacgactagcattg 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66324 TTTATTTTCATTCAATTTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTG 66265

Qy     361 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtagaatatt 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66264 GAAAATACATATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAGAATATT 66205

Qy     421 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66204 TTTGGATAGCGTTGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTAT 66145

Qy     481 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66144 TTTCATTTTTTGCTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTG 66085

Qy     541 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 600
```

```
Db  66084 AAAGAGACAAAAATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTA 66025

Qy    601 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 660
Db  66024 TTCCATCTCTCAAGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATT 65965

Qy    661 ttatcttaaattgtgatttttaaacgagctattatggtacggaacttttttttaatgaggaa 720
Db  65964 TTATCTTAAATTGTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTAATGAGGAA 65905

Qy    721 tttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgat 780
Db  65904 TTTCATGATGATTTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGAT 65845

Qy    781 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 840
Db  65844 GTGCCAGCTAAAATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTAT 65785

Qy    841 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 900
Db  65784 ATTCTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCT 65725

Qy    901 tccttttctcactctttctctcancactcacctctgtattcattctgttgtttgggatag 960
Db  65724 TCCTTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAG 65665

Qy    961 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 1020
Db  65664 AAAAATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGA 65605

Qy   1021 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 1080
Db  65604 CTCCAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTAGTTCAAAAGGAAA 65545

Qy   1081 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 1140
Db  65544 AGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAATCAACTTTTAAGA 65485

Qy   1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
Db  65484 TAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGC 65425

Qy   1201 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
Db  65424 ATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTC 65365

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 1320
Db  65364 TTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATT 65305

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctg 1380
Db  65304 ATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTG 65245

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
Db  65244 AGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATT 65185

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
Db  65184 AGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACC-GTCTTGAGACCTGTGAACT 65126

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccag 1560
Db  65125 GCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGAATATCTGCCATCCAG 65066

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttt 1620
Db  65065 TGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTT 65006

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc-nn 1679
```

```
Db  65005 GCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGCAAA 64946

Qy   1680 nnnnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagcttgc 1739
Db  64945 AAAAAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGCTTGC 64886

Qy   1740 tgactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaacta 1799
Db  64885 TGACTGTCCAAGGCACAAGAGAAAATTGAGGAATTGAAATGCAACCTGAGTATCAAACTA 64826

Qy   1800 aatattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaat 1859
Db  64825 AATATTCTAATCAAAGGTAGGTACTGTTAGGTGGAATTCTATCAGCAGGCAACTGCAAAT 64766

Qy   1860 gagaagaagatagaaggacgcccgtcgggactttggagggcagtgttattttcccaaaga 1919
Db  64765 GAGAAGAAGATAGAAGGACGCCCGTCGGGACTTTGGAGGGCAGTGTTATTTTCCCAAAGA 64706

Qy   1920 aagacggccaagggcagaggcatggattctttgcagagcacttcctttttggtttttcagt 1979
Db  64705 AAGACGGCCAAGGGCAGAGGCATGGATTCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGT 64646

Qy   1980 actgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagca 2039
Db  64645 ACTGTTTCATAGACAGTGGGCTCACATGTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCA 64586

Qy   2040 tcccagtt-attgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagt 2098
Db  64585 TCCCAGTTAATTGCAGTAATTAGAACTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGT 64526

Qy   2099 atgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatg 2158
Db  64525 ATGTCACATGAAGAAAATGTGAAATTCAAGAGTAATCCACACGTGAGAAACTAGACAATG 64466

Qy   2159 tacattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacac 2218
Db  64465 TACATTCATGTGTTCTCTTGAAAGGAAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACAC 64406

Qy   2219 cggagaaaagcaggaataacttttaccgtggaaataatgtttagctttttatgcagagaaaa 2278
Db  64405 CGGAGAAAAGCAGGAATAACTTTACCGTGGAAATAATGTTTAGCTTTTAT-CAGAGAAAA 64347

Qy   2279 ttgtccttcctagagcatagagtcccaaaactcaattctggttttcccc 2327
Db  64346 TTGTCCTT-CTAGAGCATAGAGTCCCAAAACTCAATTCTGGTTTTCCCC 64299
```

```
RESULT   4
AK021511
LOCUS       AK021511     1587 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ11449 fis, clone HEMBA1001411.
ACCESSION   AK021511
VERSION     AK021511.1  GI:10432707
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
            mRNA, clone_lib:HEMBA1 clone:HEMBA1001411.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1587)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
```

```
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES              Location/Qualifiers
     source           1. .1587
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="HEMBA1001411"
                      /clone_lib="HEMBA1"
                      /dev_stage="embryo, 10 weeks"
                      /tissue_type="whole embryo, mainly head"
                      /note="cloning vector: pME18SFL3"
BASE COUNT      486 a    244 c    276 g    581 t
ORIGIN
```

Query Match 68.1%; Score 1584.4; DB 89; Length 1587;
Best Local Similarity 99.9%; Pred. No. 4.3e-236;
Matches 1585; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      91 taatttttcttagtaaaaagaataacagaatgcatcgtggcaatccttaagcaacattatc 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAATTTTCTTAGTAAAAAGAATAACAGAATGCATCGTGGCAATCCTTAAGCAACATTATC 60

Qy     151 tatgtggactgcttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgaca 210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TATGTGGACTGCTTAAATCAGCAAAACACCAGAAGTTTGGTTAACTTGGGCAATATGACA 120

Qy     211 agtattactttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaa 270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTATTACTTTTTGGGCAAAACTACTCATTAAGCAATTTCTCTAGTGTGTCGGACACAAA 180

Qy     271 taggttctttatttttggcatgtatgccttttttatttttcattcaatttttttttttttctc 330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TAGGTTCTTTATTTTTGGCATGTATGCCTTTTTATTTTCATTCAATTTTTTTTTTTTCTC 240

Qy     331 agacagacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatt 390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATATCACTATTCTTGGAATATT 300

Qy     391 tatggtcagtctactttttagtagaatatttttggatagcgttgacacgatagatctttat 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TATGGTCAGTCTACTTTTTAGTAGAATATTTTTGGATAGCGTTGACACGATAGATCTTAT 360

Qy     451 tccatacttctttattattgataatttttattttcatttttttgctttcattattatacata 510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTGCTTTCATTATTATACATA 420

Qy     511 ttttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaa 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAAATTTATTATAACACTAAA 480

Qy     571 cactccttttttgacatattaaagcctttattccatctctcaagatatattataaaattt 630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCAAGATATATTATAAAATTT 540

Qy     631 atttttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagcta 690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATTTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATTGTGATTTTAAGCGAGCTA 600

Qy     691 ttatggtacggaacttttttttaatgaggaatttcatgatgatttaggaatttttctctctt 750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTATGGTACGGAACTTTTTTTTAATGAGGAATTTCATGATGATTTAGGAATTTTCTCTCTT 660

Qy     751 ggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccatttaaaa 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGAAAAGGCTTCCCCTGTGATGAAAATGATGTGCCAGCTAAAATTGTGTGCCATTTAAAA 720

Qy     811 actgaaaatattttaaaattatttgtctatattctaaattgagctttggatcaaacttta 870
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACTGAAAATATTTTAAAATTATTTGTCTATATTCTAAATTGAGCTTTGGATCAAACTTTA 780

Qy     871 ggccaggaccagctcatgcgttctcattcttccttttctcactctttctctcancactca 930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     781 GGCCAGGACCAGCTCATGCGTTCTCATTCTTCCTTTTCTCACTCTTTCTCTCATCACTCA 840

Qy     931 cctctgtattcattctgttgtttgggatagaaaaatcataaagagccaacccatctcaga 990
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CCTCTGTATTCATTCTGTTGTTTGGGATAGAAAAATCATAAAGAGCCAACCCATCTCAGA 900

Qy     991 acgttgtggattgagagagacactacatgactccaagtatatgagaaaaggacagagctc 1050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ACGTTGTGGATTGAGAGAGACACTACATGACTCCAAGTATATGAGAAAAGGACAGAGCTC 960

Qy    1051 taattgataactctgtagttcaaaaggaaaagagtatgcccaattctctctacatgacat 1110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TAATTGATAACTCTGTAGTTCAAAAGGAAAAGAGTATGCCCAATTCTCTCTACATGACAT 1020

Qy    1111 attgagatttttttttaatcaacttttaagatagtgatgttctgttctaaactgttctgtt 1170
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ATTGAGATTTTTTTTTAATCAACTTTTAAGATAGTGATGTTCTGTTCTAAACTGTTCTGTT 1080

Qy    1171 ttagtgaaggtagatttttataaaaacaagcatggggattcttttctaaggtaatattaat 1230
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTAGTGAAGGTAGATTTTTATAAAACAAGCATGGGGATTCTTTTCTAAGGTAATATTAAT 1140

Qy    1231 gagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcacattatgt 1290
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GAGAAGGGAAAAAAGTATCTTTAACAGCTCTTTGTTGAAGCCTGTGGTAGCACATTATGT 1200

Qy    1291 ttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctccaaaaat 1350
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTATAATTGCACATGTGCACATAATCTATTATGATCCAATGCAAATACAGCTCCAAAAAT 1260

Qy    1351 attaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataaactgaag 1410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACATTTTTCTTAATAAACTGAAG 1320

Qy    1411 agtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaaaacatca 1470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTTGTGTGGCTGCAAAACATCA 1380

Qy    1471 caaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaattaatgc 1530
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTTTAGTAAATAAAATTAATGC 1440

Qy    1531 atttctagaggggggaatatctgccatccagtggtggaaatgtggagtaaagaagctggtg 1590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTG 1500

Qy    1591 gtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaactttagc 1650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGC 1560

Qy    1651 tactgtctttggtttgagagccatggc 1677
           |||||||||||||||||||||||||||
Db    1561 TACTGTCTTTGGTTTGAGAGCCATGGC 1587
```

```
RESULT   5
AX067327
LOCUS       AX067327      674 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 31 from Patent WO0078960.
ACCESSION   AX067327
VERSION     AX067327.1  GI:12544951
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Yuqiu,J. and Mitcham,J.L.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0078960-A 31 28-DEC-2000;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      222 a    110 c    128 g    213 t      1 others
ORIGIN


  Query Match             10.3%;  Score 240;  DB 10;  Length 674;
  Best Local Similarity   99.2%;  Pred. No. 4.1e-28;
  Matches  262;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy    2064 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 2123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTTCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGTCACATGAAGAAAATGTGAAAT 60

Qy    2124 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 2183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCAAGAGTAATCCACACGTGAGAAACTAGACAATGTACATTCATGTGTTCTCTTGAAAGG 120

Qy    2184 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 2243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AAAGGGAGAGCTGTAAGCTTCACTCTGTCCTACACCGGAGAAAAGCAGGAATAACTTTAC 180

Qy    2244 cgtggaaataatgtttagcttttatgcagagaaaattgtccttcctagagcatagagtcc 2303
           |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db     181 CGTGGAAATAATGTTTAGCTTTTAT-CAGAGAAAATTGTCCTT-CTAGAGCATAGAGTCC 238

Qy    2304 caaaactcaattctggttttcccc 2327
           |||||||||||||||||||||||||
Db     239 CAAAACTCAATTCTGGTTTTCCCC 262


RESULT   6
PFMAL4P4/c
LOCUS       PFMAL4P4   224448 bp    DNA             HTG       11-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL035477
VERSION     AL035477.5  GI:5731932
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 224448)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 12, 1999 this sequence version replaced gi:5531400.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .224448
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="4"
BASE COUNT    80792 a  23258 c  21957 g  75240 t  23201 others
ORIGIN


  Query Match             3.7%;  Score 86.2;  DB 84;  Length 224448;
  Best Local Similarity   48.0%;  Pred. No. 0.00013;
  Matches  276;  Conservative    0;  Mismatches  298;  Indels    1;  Gaps    1;

Qy     275 ttctttatttttggcatgtatgccttttattttcattcaatttttttttttttctcagac 334
           ||  || ||| |   | | |    ||||| |||| || |||||  |||||||| |   ||
Db  207338 TTTATTTTTTATTTAAAGAAATTTTTTTTTTTTTTTTTTTAATTTAATTTTTTTTTTTAAC 207279

Qy     335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
           |   |   | | |   ||  | ||| || |  | ||       | ||| |
Db  207278 ATTTTTTCCCTTTTCTTTTTATTTTATGATATATATTTTTTTTTTTAACATTTTTTTTTTTT 207219

Qy     395 gtcagtctacttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
           |   | | | |||| || | |  ||||   |    ||   | |    | ||  |
Db  207218 TTTTTTTGTTATTTAATAAATTTGTTTTTATTTTATTTCTAATAAATTTTTTTTATAAT 207159

Qy     455 tacttctttattattgataatttttattttcatttttgctttcattattatacatatttt 514
           | || ||| ||||  | |||||| |||| |||||||  |   | |||| | ||| ||||
Db  207158 CATTTTTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATATTTTTCATTTTTT 207099

Qy     515 ggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaacact 574
           |  |  | |  ||    |||| | | |      |  |||| ||| || | |   | |
Db  207098 ATTCTATCAAAATTTATATTTTTATTATAATTTTTATTATTT-TTAAAATATTTTTCTCC 207040

Qy     575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
           ||||||    | ||      |||| |  || | |    || |  | |  ||| ||||
Db  207039 TTTTTTTTTTTTTTTTTTTTATTTTAATAAATTTGTTTTTATTTTATTTCTAATAAATTT 206980

Qy     635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
           ||||  ||    |||  |    |   |    ||   || ||| |||       | ||  ||
Db  206979 TTTTTTTTATTCATTATTTTTATTAATATATTTTTTTTCTGAATTTTTTTTTGATAAAAT 206920

Qy     695 ggtacggaacttttttaatgaggaatttcatgatgatttaggaattttctctcttggaa 754
           |      || |||||  |    | |  |  | ||||    || |   | ||  ||
Db  206919 TTTTTATTTTTTATTTTATCAAAATTTATTTTTTTTATTTTAATTTTTATTATTTTTAAAA 206860

Qy     755 aaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccatttaaaaactg 814
           |   ||| ||| |  |      | |||       | ||    || |   ||  |   ||
Db  206859 TATTTTTCTCCTTTTTTTTTTTTTTTGTTTTTATTTAATAAATTTTTTTTTTATATTTCTT 206800

Qy     815 aaaatattttaaaattatttgtctatattctaaat 849
           | ||||    || |||   | | ||||  ||
Db  206799 TTTTTTTTTTTCTTTTTTTTCATTTTTTTCTTTAT 206765


RESULT   7
CNS01JRG
LOCUS       CNS01JRG      879 bp    DNA             STS       17-FEB-2000
DEFINITION  Anopheles gambiae STS T7 end of clone 14D07 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            sequence tagged site.
ACCESSION   AL147405
VERSION     AL147405.1  GI:7005551
KEYWORDS    STS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 879)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="14D07"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT       66 a     38 c      1 g     670 t     104 others
ORIGIN


  Query Match             3.5%;  Score 82.6;  DB 53;  Length 879;
  Best Local Similarity   43.0%;  Pred. No. 0.00089;
  Matches  246;  Conservative   26;  Mismatches  295;  Indels    5;  Gaps    1;

Qy     268 aaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttttttt 327
           |:| |   | ||| |||||    | | |   ||||| ||||  ||   ||||||||||||
Db       1 AWAAATAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 60

Qy     328 ctcagacagacatagtagtaacgactagcattggaaaatacatatcactattcttggaat 387
           | .       |  |  |||  : |    ||            |   | ||  |  : |
Db      61 TTTNATANNNNTTNTTNNTAAKAVKTGNNTTTNNNNNNNNNNNTTTTTTTTTTNTATRTT 120

Qy     388 atttatggtcagtctactttttagtagaatatttttggatagcgttgacacgatagatct 447
           ||| |  |   | |  |||||  |   || |||||   |     ||       |   | |
Db     121 TTTTTTTATTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 180

Qy     448 tattccatacttctttattattgataatttttattttcatttttgctttcattattatac 507
           | ||   |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |
Db     181 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240

Qy     508 atattttggtggagaagaggttgggctttttgaaagagacaaaaatttattataacact 567
           | ||||  |            ||   ||||||             ||| || |    |
Db     241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300

Qy     568 aaacactccttttttgacatattaaagcctttattccatctctcaagatatattataaaa 627
           |  ||||||    | ||       ||| ||   | | |      | | || |
Db     301 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTT 360

Qy     628 tttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgag 687
           ||| |||||||  |||    ||| |   || |||| | ||   |  |  |||:
Db     361 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTNTTTTTTTHTTTTTTT 420

Qy     688 ctattatggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttctct 747
           | ::    ||     :| ||||| :: :  :|||  | ||  |||     ||||  | |
Db     421 TTTYHCAAMTATMCTTMTCTTTTTTTVWAWWTMATTTTTTHATTTTTT-----TTTTTTTT 475

Qy     748 cttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccattta 807
           :|        |||  |: |  :  :: |  | | ::| :  :  || | |    |:||
Db     476 MTMWATCTYTTTTTCTWCYMTTTHCTMWTTTTTTTTYYATMMTMTTTTTTTTCATTTYTA 535

Qy     808 aaaactgaaaatattttaaaattatttgtcta 839
           | |: |     | | ||: : || ||| ||::
Db     536 ACAMTTTTTTTTTTMTTMTMCTTCTTTYTCYM 567


RESULT   8
AC023371
LOCUS       AC023371    53932 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023371
VERSION     AC023371.2  GI:9123990
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 53932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21D18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 53932)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6970502.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3985
                Center clone name: 21_D_18
            ---------------.
            * NOTE: This record contains 55 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      869: contig of 869 bp in length
            *     870 969: gap of      100 bp
            *     970     1815: contig of 846 bp in length
            *    1816 1915: gap of      100 bp
            *    1916     2797: contig of 882 bp in length
            *    2798 2897: gap of      100 bp
            *    2898     3768: contig of 871 bp in length
            *    3769 3868: gap of      100 bp
            *    3869     4769: contig of 901 bp in length
            *    4770 4869: gap of      100 bp
            *    4870     5739: contig of 870 bp in length
            *    5740 5839: gap of      100 bp
            *    5840     6703: contig of 864 bp in length
            *    6704 6803: gap of      100 bp
            *    6804     7675: contig of 872 bp in length
```

```
*     7676 7775: gap of      100 bp
*     7776    8671: contig of 896 bp in length
*     8672 8771: gap of      100 bp
*     8772    9639: contig of 868 bp in length
*     9640 9739: gap of      100 bp
*     9740   10613: contig of 874 bp in length
*    10614 10713: gap of      100 bp
*    10714   11572: contig of 859 bp in length
*    11573 11672: gap of      100 bp
*    11673   12528: contig of 856 bp in length
*    12529 12628: gap of      100 bp
*    12629   13488: contig of 860 bp in length
*    13489 13588: gap of      100 bp
*    13589   14469: contig of 881 bp in length
*    14470 14569: gap of      100 bp
*    14570   15452: contig of 883 bp in length
*    15453 15552: gap of      100 bp
*    15553   16429: contig of 877 bp in length
*    16430 16529: gap of      100 bp
*    16530   17401: contig of 872 bp in length
*    17402 17501: gap of      100 bp
*    17502   18387: contig of 886 bp in length
*    18388 18487: gap of      100 bp
*    18488   19361: contig of 874 bp in length
*    19362 19461: gap of      100 bp
*    19462   20326: contig of 865 bp in length
*    20327 20426: gap of      100 bp
*    20427   21291: contig of 865 bp in length
*    21292 21391: gap of      100 bp
*    21392   22281: contig of 890 bp in length
*    22282 22381: gap of      100 bp
*    22382   23271: contig of 890 bp in length
*    23272 23371: gap of      100 bp
*    23372   24249: contig of 878 bp in length
*    24250 24349: gap of      100 bp
*    24350   25213: contig of 864 bp in length
*    25214 25313: gap of      100 bp
*    25314   26162: contig of 849 bp in length
*    26163 26262: gap of      100 bp
*    26263   27122: contig of 860 bp in length
*    27123 27222: gap of      100 bp
*    27223   28086: contig of 864 bp in length
*    28087 28186: gap of      100 bp
*    28187   29073: contig of 887 bp in length
*    29074 29173: gap of      100 bp
*    29174   30055: contig of 882 bp in length
*    30056 30155: gap of      100 bp
*    30156   31136: contig of 981 bp in length
*    31137 31236: gap of      100 bp
*    31237   32116: contig of 880 bp in length
*    32117 32216: gap of      100 bp
*    32217   33100: contig of 884 bp in length
*    33101 33200: gap of      100 bp
*    33201   34070: contig of 870 bp in length
*    34071 34170: gap of      100 bp
*    34171   35169: contig of 999 bp in length
*    35170 35269: gap of      100 bp
*    35270   36164: contig of 895 bp in length
*    36165 36264: gap of      100 bp
*    36265   37135: contig of 871 bp in length
*    37136 37235: gap of      100 bp
*    37236   38163: contig of 928 bp in length
*    38164 38263: gap of      100 bp
*    38264   39164: contig of 901 bp in length
*    39165 39264: gap of      100 bp
*    39265   40131: contig of 867 bp in length
*    40132 40231: gap of      100 bp
*    40232   41124: contig of 893 bp in length
*    41125 41224: gap of      100 bp
*    41225   42122: contig of 898 bp in length
*    42123 42222: gap of      100 bp
*    42223   43129: contig of 907 bp in length
*    43130 43229: gap of      100 bp
*    43230   44109: contig of 880 bp in length
*    44110 44209: gap of      100 bp
*    44210   45090: contig of 881 bp in length
*    45091 45190: gap of      100 bp
*    45191   46090: contig of 900 bp in length
*    46091 46190: gap of      100 bp
*    46191   47072: contig of 882 bp in length
*    47073 47172: gap of      100 bp
*    47173   48053: contig of 881 bp in length
*    48054 48153: gap of      100 bp
*    48154   49033: contig of 880 bp in length
*    49034 49133: gap of      100 bp
*    49134   50009: contig of 876 bp in length
*    50010 50109: gap of      100 bp
*    50110   50981: contig of 872 bp in length
*    50982 51081: gap of      100 bp
*    51082   51957: contig of 876 bp in length
*    51958 52057: gap of      100 bp
*    52058   52945: contig of 888 bp in length
*    52946 53045: gap of      100 bp
*    53046   53932: contig of 887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .53932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-21D18"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    14211 a   9672 c   9208 g  15109 t   5732 others
ORIGIN


  Query Match              3.4%;  Score 80.2;  DB 68;  Length 53932;
  Best Local Similarity   48.1%;  Pred. No. 0.0013;
  Matches  269;  Conservative    0;  Mismatches  285;  Indels    5;  Gaps    2;

Qy     311 ttcaatttttttttttctcagacagacatagtagtaacgactagcattggaaaatacat 370
           || | | ||||||| || | | | |   ||| |  | |  |  |   ||  | | |   |
Db   29459 TTTATTNTTTTTTTATTTTTATAAANTTATATTTTTTAAAAAAAAATTTTTATATTTTTT 29518

Qy     371 atcactattcttggaatatttatggtcagtctacttttagtagaatatttttggatagc 430
            |    || ||  | ||||| |  |   |    |||||  ||  || ||||   ||
Db   29519 TTTTTNTTTTTTTTATTATTTTTTTTTTTTTNTATTTTTTNTAATATTTTTTATTAT--- 29575

Qy     431 gttgacacgatagatcttattccatacttctttattattgataatttttattttcatttt 490
           ||   |   |   | || |   |||  | | ||   || || ||  | ||||  |||||
Db   29576 -TTTTTATTTTTTTTTTTTTNATATAAATTTATAAATTTAATTATAATTTTTTTTTTTTT 29634

Qy     491 tgctttcattattatacatattttggtggagaagaggttgggctttttgaaagagacaa 550
           |  ||| ||| || ||  | ||||     |       ||    | || | | |   |
Db   29635 TTTTTTTATTTTTTTAATTTTTTTATATTATTTATTTTTTNTATATTATAATATTTAATT 29694

Qy     551 aaatttattataacactaaacactccttttttgacatattaaagcctttattccatctct 610
              ||| | ||       || | |  ||||||   ||||||      |||||   |   |
Db   29695 TTTTTTTTAATTTATTATAAAATTTTTTTTTTTTTATATTATTTTTATTATTATTTNATT 29754

Qy     611 caagatatattataaaatttatttttttaatttaagatttctgaattattttatcttaaa 670
             || | ||  ||||||| |||| ||  |||      | |   || || | | ||| |
Db   29755 TTTAATTTTTTTAAAAATTTTTTTTNTTTTTTTTTTTAATTTTTTTTTTTATTTTTATA 29814

Qy     671 ttgtgattttaaacgagctattatggtacggaacttttttaatgaggaatttcatgatg 730
           || | |||||       | | ||   |    | ||||||||      ||   | ||
Db   29815 TT-TTATTTTTTTAATTTTTTAATTAAATTTTATTTTTTTTTATATTTATTTTNTNTTATT 29873

Qy     731 atttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagcta 790
           |||||  | || | |       |||     | | |  |   ||| ||     |  ||
Db   29874 ATTTATTATTTATTTTAAAATTAAATTATATATATTATTTTATTAATTATAATTTAATTA 29933

Qy     791 aaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatattctaaatt 850
           |    | |   ||| |   |     |   || |||||  ||||| | || | | | | ||
Db   29934 TATAAATATTTTATTAATTTATNATATTAATATTAAATATATTTTTATAAAATATTATTT 29993

Qy     851 gagctttggatcaaacttt 869
               |    || ||  |||
```

```
Db  29994 TTTTTAATAATTAATTTTT 30012


RESULT  9
CNS07ANM/c
LOCUS       CNS07ANM      936 bp    DNA              STS       12-JAN-2001
DEFINITION  T7 end of clone BC0AA005E01 of library BC0AA from strain CBS 767 of
            Debaryomyces hansenii, sequence tagged site.
ACCESSION   AL436808
VERSION     AL436808.1  GI:12220221
KEYWORDS    STS.
SOURCE      Debaryomyces hansenii.
  ORGANISM  Debaryomyces hansenii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Debaryomyces.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   Lepingle,A., Casaregola,S., Neuveglise,C., Bon,E., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 14.
            Debaryomyces hansenii var. hansenii
  JOURNAL   FEBS Lett. 487 (1), 82-86 (2000)
  PUBMED    11152889
REFERENCE   2  (bases 1 to 936)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   3  (bases 1 to 936)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .936
                     /organism="Debaryomyces hansenii"
                     /strain="CBS 767"
                     /variety="hansenii"
                     /db_xref="taxon:4959"
                     /clone="BC0AA005E01"
                     /clone_lib="BC0AA"
                     /note="end : T7"
     STS             1. .936
BASE COUNT      607 a     67 c     61 g    167 t     34 others
ORIGIN


  Query Match             3.4%;  Score 79.2;  DB 54;  Length 936;
  Best Local Similarity   45.2%;  Pred. No. 0.003;
  Matches  247;  Conservative   19;  Mismatches  277;  Indels    3;  Gaps    1;

Qy    213 tattacttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaata 272
          | | ||| |||||    |  |  |  ||     |||| | || ||| |  |       |
Db    667 TGTAACTATTTGGTTTTATTTTATTTTTTTTKTATTTTTTTATTGTATTTGTTGTTTTTT 608

Qy    273 ggttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcag 332
            || ||| || ||    |  ||   ||:|||| |||  |   ||||||: :||    |
Db    607 TTTTATTTTTTATTTTTTTTAGATTTTTTKTTATATTCTATTCGTTTTTTWGWTTAAAGAT 548

Qy    333 acagacatagtagtaacgactagcattggaaaatacatatcactattcttggaatattta 392
          | |||    |   || |   |  |  |  |   ||   |||: |     |  ||
Db    547 TTCTAGATATGTTTTAGATTTTTTATTATTTGTATAATTTTTATYTTCTCTTTTATTT 488

Qy    393 tggtcagtctacttttttagtagaatatttttggatagcgttgacacgatagatcttattc 452
          | |   | || |||||   | |   | |||                  |:: | || ||
Db    487 TATTTCTTTTATTTTTTTTAAWATATTATTTACTATTGAGGTTTTTTTTWKCTTTTTTTT 428

Qy    453 catacttctttattattgataatttttattttcatttttttgctttcattattatacatatt 512
          :|: || ||| | |:|  |  | :| |||| ||||||  |||  ||:|| |   | ||
Db    427 TWTWTTTTTTTTTTAAWTTTTTTTAWTGTTTTAGTTTTTTTTTTTTTTTTTWTTTTTTTTTT 368

Qy    513 ttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaaca 572
          ||  |          ||    |||||      |||  |   ||| || | || ||  | |
Db    367 TTTTTWTTTTTTTTTTTTTTTTTTTTTCAGKGAAGATTATCCTTTTTTTTTTACCCTTKATA 308

Qy    573 ctccttttt---tgacatattaaagcctttattccatctctcaagatatattataaaatt 629
              |||||   |  | || | | |  | |:  |   |    | : || |    ||
Db    307 TATTTTTTTTACTTTCKTAATTATGATTCTTTKGTTTTCATTTTTTTTWTTTTTTTTTTT 248

Qy    630 tatttttttaatttaagatttctgaattatttttatcttaaattgtgattttaaacgagct 689
          | |||||||  ||:    |||  |  |||||||| | ||   || | |||||     :  |
Db    247 TTTTTTTTTTTTTWTTWTTTTTTATATTATTTTTTATTTTTTTTTAATTTTTTTTTWTTT 188

Qy    690 attatggtacggaacttttttttaatgaggaatttcatgatgatttaggaatttttctctct 749
          || |  | |   | |||||||  | |    |||| |  | |||      :::| | | |
Db    187 TTTTTTWTTCTATATTTTTTTTTTTTAATTTTTTTCTTTTTTTATTACTTTCWWWTTTTTTT 128

Qy    750 tggaaa 755
          :  |||
Db    127 WTTAAA 122


RESULT  10
CNS06PLB
LOCUS       CNS06PLB      1084 bp    DNA              STS       10-JAN-2001
DEFINITION  T7 end of clone AV0AA015D02 of library AV0AA from strain CBS 379 of
            Saccharomyces exiguus, sequence tagged site.
ACCESSION   AL409509
VERSION     AL409509.1  GI:12177058
KEYWORDS    STS.
SOURCE      Saccharomyces exiguus.
  ORGANISM  Saccharomyces exiguus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 6.
            Saccharomyces exiguus
  JOURNAL   FEBS Lett. 487 (1), 42-46 (2000)
  PUBMED    11152881
REFERENCE   2  (bases 1 to 1084)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   3  (bases 1 to 1084)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
```

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Saccharomyces exiguus"
                     /strain="CBS 379"
                     /db_xref="taxon:34358"
                     /clone="AV0AA015D02"
                     /clone_lib="AV0AA"
                     /note="end : T7"
     misc_feature    complement(<17. .>979)
                     /note="similar to Saccharomyces cerevisiae ORF YMR304w [
                     UBP15 ; similarity to human ubiquitin-specific protease ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      214 a    114 c     92 g    630 t     34 others
ORIGIN
```

```
  Query Match             3.4%;  Score 78.6;  DB 53;  Length 1084;
  Best Local Similarity   44.8%;  Pred. No. 0.0036;
  Matches  230;  Conservative   24;  Mismatches  258;  Indels    1;  Gaps    1;

Qy    238 attaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgc 297
          |:||  ||| || : :| | |:||| | | || :|  || | || || | :   :
Db     83 AKTATCCAAATTTKAYATTATKTCGAATAGAACKATTTTATCTAATTCTTCYTCAYTATW 142

Qy    298 ctttttatttttcattcaatttttttttttttctcagacagacatagtagtaacgactagca 357
          |||||::|||  || :|||:|||| :||| |       : :|  |: ||   : |   |
Db    143 ATTTTTWWTTTATTTAWATTWTTTTCYTTTTTGTTTWTTWAWTTTTW-TAYTTWTTTTAA 201

Qy    358 ttggaaaatacatatcactattcttggaatatttatggtcagtctactttttagtagaat 417
          ||    | |:  | |   |||| ::   |::||| |  |   | |  ||||| |    |
Db    202 TTTTTTATTWTTTTTTTTTATTTWWWYTAWWTTTTTTTTTTTTTTTTTTTTTTTTTTTT 261

Qy    418 atttttggatagcgttgacacgatagatcttattccatacttctttattattgataattt 477
          |||||  |   ||       |   | || ||   |  || ||| || ||  |  |||
Db    262 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 321

Qy    478 tattttcatttttgctttcattattatacatattttggtggagaagaggttgggctttt 537
          | |||   ||||||  ||| |||:||:|     ||||  |          ||    ||||
Db    322 TTTTTATTTTTTTTTATTTTATTWTTWTTTTCTTTTTTTTTTTTTTTTTTTTTTCTTTTT 381

Qy    538 ttgaaagagacaaaaatttattataacactaaacactcctttttgacatattaaagcct 597
          ||              ||| || |     |  |   |  ||||||    | ||      |
Db    382 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTT 441

Qy    598 ttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaatt 657
          | |||   | | |     | | || |    ||| |||||||  |||    ||| |   ||
Db    442 TAATTTTYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 501

Qy    658 attttatcttaaattgtgattttaaacgagctattatggtacggaacttttttaatgag 717
          || | | ||   || |  |||         | || |   |     | |||||||  |
Db    502 TTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTT 561

Qy    718 gaatttcatgatgatttaggaattttctctctt 750
             |||     |  |:|     |||| | | ||
Db    562 TTTTTTTTATTTTTTYTTTTTTTTTTTTTTTT 594
```

```
RESULT  11
CNS06KHQ
LOCUS       CNS06KHQ     1169 bp    DNA             STS       10-JAN-2001
DEFINITION  T3 end of clone AT0AA005D05 of library AT0AA from strain CBS 4311
            of Saccharomyces servazzii, sequence tagged site.
ACCESSION   AL402900
VERSION     AL402900.1  GI:12162508
KEYWORDS    STS.
SOURCE      Saccharomyces servazzii.
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1169)
  AUTHORS   Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 7.
            Saccharomyces servazzii
  JOURNAL   FEBS Lett. 487 (1), 47-51 (2000)
   PUBMED   11152882
REFERENCE   2  (bases 1 to 1169)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 1169)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1169
                     /organism="Saccharomyces servazzii"
                     /strain="CBS 4311"
                     /db_xref="taxon:27293"
                     /clone="AT0AA005D05"
                     /clone_lib="AT0AA"
                     /note="end : T3"
     STS             1. .1169
                     /note="part of mitochondrial DNA"
                     /evidence=not_experimental
BASE COUNT      375 a    101 c     57 g    545 t     91 others
ORIGIN
```

```
  Query Match             3.3%;  Score 76.8;  DB 53;  Length 1169;
  Best Local Similarity   43.7%;  Pred. No. 0.0068;
  Matches  211;  Conservative   25;  Mismatches  244;  Indels    3;  Gaps    1;

Qy    204 tatgacaagtattacttttgggcaaaactactcattaagcaatttctctagtgtgtcgg 263
          :|| |  |  ||   | | |    | |  |    ||  |  ||||  | || | | |
Db    646 WATAATTAANATATATATTTATATATATTTATATATATAAATTNTTATATTATATAAA 705

Qy    264 acacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttt 323
          |    ||:|  |  ||| |  |    || | |    |  |  |   ||| | :|||||||
Db    706 AATATAAWATATATTTTTTAATATATATNTTTTTTATAATTATAAAATTTATWTTTTTTT 765

Qy    324 ttttctcagacagacatagtagtaacgactagcattggaaaatacatatcactattcttg 383
          ||:| |   | | | ||| || :|:    ||   ||     |  | ||||   |     |
Db    766 TTWTTTTTTATATAAATAATATWAWTATTTATATTTATNTATAATATATATATNANTNTA 825

Qy    384 gaatatttatggtcagtctactttttagtagaatatttttggatagcgttgacacgatag 443
          |||  |   |   | || :|  :|  : ||: |:||   ::    || |     |
Db    826 TTTTATNTTATTTTNTTTTAWWTAAWAAAWTAAWTTWTTNTAWWTTATTTWAATTTTTTT 885
```

```
Qy     444 atcttattccatacttctttattattgat---aattttattttcatttttgctttcatt 500
           | || ||    |  || :||| |    |    |||  : ||||  || |||  ||  |||
Db     886 TTTTTTTTTTTTTTTTTTWTTANTTANTAATTAAATAAWTTTTTTTTTATTTTWTTATATT 945

Qy     501 attatacatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttatta 560
           ||| |: || ||||  |  |   |  ||    ||||||   :      || :::| |
Db     946 ATTTTWTATTTTTTTTATATATTTTATTTTTATWTTTTTTTTTWNTWTTTAATWWWTTTAT 1005

Qy     561 taacactaaacactccttttttgacatattaaagcctttattccatctctcaagatatat 620
           : : |   |  | |  ||||      | |:  |    |  ||   : | |  |  | |||
Db    1006 WTWTAWATATAATTTATTTTAATNWNTTTWTTATATATATTTATTWTTNTATATTTTTAT 1065

Qy     621 tataaaatttatttttttaatttaagatttctgaattattttatcttaaattgtgatttt 680
           | | | |||||||| |||  ||||   ||| | |:|| || ||| |  | |    |||
Db    1066 TTTTATATTTATTTATTTTTTTTATTTTTTTTNAWTTTTTATATATATATTATATTTTTA 1125

Qy     681 aaa 683
           |||
Db    1126 TAA 1128
```

```
RESULT  12
CNS07ABZ
LOCUS       CNS07ABZ      897 bp    DNA              STS       12-JAN-2001
DEFINITION  T7 end of clone BC0AA003A12 of library BC0AA from strain CBS 767 of
            Debaryomyces hansenii, sequence tagged site.
ACCESSION   AL436389
VERSION     AL436389.1  GI:12219802
KEYWORDS    STS.
SOURCE      Debaryomyces hansenii.
  ORGANISM  Debaryomyces hansenii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Debaryomyces.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Lepingle,A., Casaregola,S., Neuveglise,C., Bon,E., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 14.
            Debaryomyces hansenii var. hansenii
  JOURNAL   FEBS Lett. 487 (1), 82-86 (2000)
  PUBMED    11152889
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   3  (bases 1 to 897)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Debaryomyces hansenii"
                     /strain="CBS 767"
                     /variety="hansenii"
                     /db_xref="taxon:4959"
                     /clone="BC0AA003A12"
                     /clone_lib="BC0AA"
                     /note="end : T7"
     STS             1. .897
                     /note="part of mitochondrial DNA"
                     /evidence=not_experimental
BASE COUNT      148 a      70 c      47 g     590 t      42 others
ORIGIN
```

```
  Query Match             3.3%;  Score 76.4;  DB 54;  Length 897;
  Best Local Similarity   45.0%;  Pred. No. 0.008;
  Matches  215;  Conservative   16;  Mismatches  247;  Indels    0;  Gaps    0;

Qy     203 atatgacaagtattacttttttgggcaaaactactcattaagcaatttctctagtgtgtcg 262
           |||| |  | |||   |:||           |   || |  | | |   || | |  :
Db     200 ATATAATTATTATATTTKTTATACACCGKGCAGATATCACTTACTATACTTACTATTAYT 259

Qy     263 gacacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttt 322
           :   |||  || ||| |||||   | | |   ||:||||| |  ||   |||||||
Db     260 WTTWTGTATAATTTATTTTTTTTTTTWTTWTTTTTTTTTWTTATTATTTTTTTTTTTTTTT 319

Qy     323 tttttctcagacagacatagtagtaacgactagcattggaaaatacatatcactattctt 382
           ||||| |          ||  |  |       |    |       |   | |   | || ||
Db     320 TTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTT 379

Qy     383 ggaatatttatggtcagtctacttttagtagaatatttttggatagcgttgacacgata 442
           | ||| :  | : | |  |||||  |    | |:|||   ||   ||      :|
Db     380 TTTTTTTTTWTTTTWTTTTTTTTTTTTTTTTTTTTTWTTTTTTTATTWTTTTTTTTWTT 439

Qy     443 gatcttattccatacttctttattattgataattttatttttcatttttgctttcattat 502
           | |: ||   |   | ||| || || ||  |||| |||| |||||||  ||| :||||
Db     440 TTTTWTTTTTTTTTATTTTTTTTTTTAATTTTTTTTTTTTTTATTTTTTTTTTTTWTTAT 499

Qy     503 tatacatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattata 562
           | :   |  ||| | :      ||    ||||||     |        ||||||:|
Db     500 TTWTTTTTATTTATTTTWTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTATTWTT 559

Qy     563 acactaaacactccttttttgacatattaaagcctttattccatctctcaagatatatta 622
           | |      |  ||:||  |  | |:      ||| ||   | | | :   | | ||
Db     560 TWATTTTTTTTTTTTTTWTTATATTTTTWTTTTTWTTTTTTTTTTTWTTTAWTATTTTTTTT 619

Qy     623 taaaatttatttttttaatttaagatttctgaattattttatcttaaattgtgatttt 680
           |    ||| |||||||  |||   |||| |  ||| || | | ||   || |  ||||
Db     620 TTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTATTTTTATTTTTTTTTTTTTTTTTT 677
```

```
RESULT  13
AC023466
LOCUS       AC023466    60604 bp    DNA              HTG       13-JUL-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-661A23 map 7, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC023466
VERSION     AC023466.2  GI:9168545
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60604)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 7, clone RP11-661A23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60604)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
```

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R., Liew,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jul 13, 2000 this sequence version replaced gi:6970686.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L6844
Center clone name: 661_A_23

* NOTE: This record contains 67 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 849: contig of 849 bp in length
* 850 949: gap of 100 bp
* 950 1760: contig of 811 bp in length
* 1761 1860: gap of 100 bp
* 1861 2678: contig of 818 bp in length
* 2679 2778: gap of 100 bp
* 2779 3603: contig of 825 bp in length
* 3604 3703: gap of 100 bp
* 3704 4497: contig of 794 bp in length
* 4498 4597: gap of 100 bp
* 4598 5410: contig of 813 bp in length
* 5411 5510: gap of 100 bp
* 5511 6309: contig of 799 bp in length
* 6310 6409: gap of 100 bp
* 6410 7184: contig of 775 bp in length
* 7185 7284: gap of 100 bp
* 7285 8075: contig of 791 bp in length
* 8076 8175: gap of 100 bp
* 8176 8987: contig of 812 bp in length
* 8988 9087: gap of 100 bp
* 9088 9902: contig of 815 bp in length
* 9903 10002: gap of 100 bp
* 10003 10822: contig of 820 bp in length
* 10823 10922: gap of 100 bp
* 10923 11722: contig of 800 bp in length
* 11723 11822: gap of 100 bp
* 11823 12582: contig of 760 bp in length
* 12583 12682: gap of 100 bp
* 12683 13489: contig of 807 bp in length
* 13490 13589: gap of 100 bp
* 13590 14355: contig of 766 bp in length
* 14356 14455: gap of 100 bp
* 14456 15239: contig of 784 bp in length
* 15240 15339: gap of 100 bp
* 15340 16151: contig of 812 bp in length
* 16152 16251: gap of 100 bp
* 16252 17044: contig of 793 bp in length
* 17045 17144: gap of 100 bp
* 17145 17948: contig of 804 bp in length
* 17949 18048: gap of 100 bp
* 18049 18855: contig of 807 bp in length
* 18856 18955: gap of 100 bp
* 18956 19759: contig of 804 bp in length
* 19760 19859: gap of 100 bp
* 19860 20664: contig of 805 bp in length
* 20665 20764: gap of 100 bp
* 20765 21567: contig of 803 bp in length
* 21568 21667: gap of 100 bp
* 21668 22483: contig of 816 bp in length
* 22484 22583: gap of 100 bp
* 22584 23380: contig of 797 bp in length
* 23381 23480: gap of 100 bp
* 23481 24291: contig of 811 bp in length
* 24292 24391: gap of 100 bp
* 24392 25190: contig of 799 bp in length
* 25191 25290: gap of 100 bp
* 25291 26110: contig of 820 bp in length
* 26111 26210: gap of 100 bp
* 26211 27023: contig of 813 bp in length
* 27024 27123: gap of 100 bp
* 27124 27914: contig of 791 bp in length
* 27915 28014: gap of 100 bp
* 28015 28806: contig of 792 bp in length
* 28807 28906: gap of 100 bp
* 28907 29718: contig of 812 bp in length
* 29719 29818: gap of 100 bp
* 29819 30609: contig of 791 bp in length
* 30610 30709: gap of 100 bp
* 30710 31489: contig of 780 bp in length
* 31490 31589: gap of 100 bp
* 31590 32396: contig of 807 bp in length
* 32397 32496: gap of 100 bp
* 32497 33283: contig of 787 bp in length
* 33284 33383: gap of 100 bp
* 33384 34154: contig of 771 bp in length
* 34155 34254: gap of 100 bp
* 34255 35100: contig of 846 bp in length
* 35101 35200: gap of 100 bp
* 35201 36031: contig of 831 bp in length
* 36032 36131: gap of 100 bp
* 36132 36954: contig of 823 bp in length
* 36955 37054: gap of 100 bp
* 37055 37959: contig of 905 bp in length
* 37960 38059: gap of 100 bp
* 38060 38854: contig of 795 bp in length
* 38855 38954: gap of 100 bp
* 38955 39764: contig of 810 bp in length
* 39765 39864: gap of 100 bp
* 39865 40647: contig of 783 bp in length
* 40648 40747: gap of 100 bp
* 40748 41559: contig of 812 bp in length
* 41560 41659: gap of 100 bp
* 41660 42448: contig of 789 bp in length
* 42449 42548: gap of 100 bp
* 42549 43343: contig of 795 bp in length
* 43344 43443: gap of 100 bp
* 43444 44234: contig of 791 bp in length
* 44235 44334: gap of 100 bp
* 44335 45143: contig of 809 bp in length
* 45144 45243: gap of 100 bp
* 45244 46041: contig of 798 bp in length
* 46042 46141: gap of 100 bp
* 46142 46937: contig of 796 bp in length
* 46938 47037: gap of 100 bp
* 47038 47830: contig of 793 bp in length
* 47831 47930: gap of 100 bp

```
            *    47931    48734: contig of 804 bp in length
            *    48735 48834: gap of      100 bp
            *    48835    49633: contig of 799 bp in length
            *    49634 49733: gap of      100 bp
            *    49734    50517: contig of 784 bp in length
            *    50518 50617: gap of      100 bp
            *    50618    51470: contig of 853 bp in length
            *    51471 51570: gap of      100 bp
            *    51571    52387: contig of 817 bp in length
            *    52388 52487: gap of      100 bp
            *    52488    53299: contig of 812 bp in length
            *    53300 53399: gap of      100 bp
            *    53400    54204: contig of 805 bp in length
            *    54205 54304: gap of      100 bp
            *    54305    55129: contig of 825 bp in length
            *    55130 55229: gap of      100 bp
            *    55230    56063: contig of 834 bp in length
            *    56064 56163: gap of      100 bp
            *    56164    56939: contig of 776 bp in length
            *    56940 57039: gap of      100 bp
            *    57040    57851: contig of 812 bp in length
            *    57852 57951: gap of      100 bp
            *    57952    58782: contig of 831 bp in length
            *    58783 58882: gap of      100 bp
            *    58883    59712: contig of 830 bp in length
            *    59713 59812: gap of      100 bp
            *    59813    60604: contig of 792 bp in length.
FEATURES             Location/Qualifiers
     source          1. .60604
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

  Query Match              3.3%;  Score 76.2;  DB 68;  Length 60604;
  Best Local Similarity   47.4%;  Pred. No. 0.0052;
  Matches  241;  Conservative    0;  Mismatches  264;  Indels    3;  Gaps    1;

Qy     236 tcattaagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtat 295
           | |||    |||   | |  | | |   | |    ||   | ||| |||||   || | |
Db   37394 TTATTTTTTAATAAATATTATTTTTTTTATATTTTATTATTTTTTTTTTTTTTATTTTT 37453

Qy     296 gccttttttattttcattcaatttttttttttctcagacagacatagtagtaacgactag 355
           ||||||| |    |   |||||| ||||| | | |   |  || |   ||    |
Db   37454 AAATTTTTATATAATATTTTTTTTTTATTTTTATTATAATAATTTAATTTAAAATTTAAT 37513

Qy     356 cattggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaga 415
             ||  |||| | | |  |   ||     ||||  ||   | | | || ||  || |
Db   37514 ATTTTTAAAAAAAAAAAAATATTTTNANTAATAAATAAAATTATTATATTTAATAATTTT 37573

Qy     416 atatttttggatagcgttgacacgatagatcttattccatacttctttattattgataat 475
           | |||||    |   || | |  ||     || ||   |   | |||| || | |  ||
Db   37574 TTTTTTTTATTAATTTTTAATATAATTTTATTTTTTTATTTTATTTTTAATAATTAATAT 37633

Qy     476 tttatttttcatttttttgctttcattattatacatattttggtggagaagaggttgggctt 535
           | || ||| |   ||   |||  |||||||    || ||  |       |  ||     |
Db   37634 TATANTTTAAAAATTATTTTTNTTTATTATTTTAATATTTTTATTATTTANTTTTAAAAT 37693

Qy     536 ttttgaaagagacaaaaatttattataacactaaacactccttttttgacatattaaagc 595
             ||| |   || | || ||| ||    |  | | |  ||| || | | | ||
Db   37694 NAAAAAAATATTTAATATTTAATTTTATTTTTTTAAAATAATTTATTAATANAATATTAT 37753

Qy     596 ctttattccatctctcaagatatattataaaatttatttttttaatttaagatttctgaa 655
           | ||||   |   | ||   | || | ||| ||||     |    | | |||| | ||
Db   37754 ATATATT---TTAATTAATTNTTTTTTTTAAAATTATAAAAATTTAATTANATTTTTTAA 37810

Qy     656 ttattttatcttaaattgtgattttaaacgagctattatggtacggaacttttttaatg 715
           |   | |||     ||    || | |||      | |||  |    || ||  |  | |
Db   37811 TATNTATATAAATTATATATATATAAAATATATAAATATTTTNTTAAAATTAATAAATTT 37870

Qy     716 aggaatttcatgatgatttaggaatttt 743
             ||| | || |  | ||   | || |
Db   37871 TAAAATATAATAAAAAATTTTAATTTAT 37898
```

```
RESULT  14
AC079417/c
LOCUS       AC079417   130540 bp    DNA              HTG       01-SEP-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-102A24, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC079417
VERSION     AC079417.1  GI:9958029
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 130540)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 130540)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_102A24
            -------
            Summary Statistics
            Consensus quality: 71349 bases at least Q40
            Consensus quality: 83788 bases at least Q30
            Consensus quality: 89369 bases at least Q20
            Estimated insert size: 93300; agarose-fp estimation
            Estimated insert size: 127240; sum-of-contigs estimation
            Quality coverage:  3.69 in Q20 bases; agarose-fp estimation
            Quality coverage:  2.7 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1043: contig of 1043 bp in length
            *     1044     1143: gap of unknown length
            *     1144     2160: contig of 1017 bp in length
            *     2161     2260: gap of unknown length
            *     2261     3696: contig of 1436 bp in length
            *     3697     3796: gap of unknown length
            *     3797     4826: contig of 1030 bp in length
            *     4827     4926: gap of unknown length
            *     4927     5959: contig of 1033 bp in length
            *     5960     6059: gap of unknown length
            *     6060     7579: contig of 1520 bp in length
            *     7580     7679: gap of unknown length
            *     7680     9178: contig of 1499 bp in length
            *     9179     9278: gap of unknown length
            *     9279    10501: contig of 1223 bp in length
            *    10502    10601: gap of unknown length
            *    10602    11610: contig of 1009 bp in length
            *    11611    11710: gap of unknown length
            *    11711    12840: contig of 1130 bp in length
            *    12841    12940: gap of unknown length
            *    12941    14262: contig of 1322 bp in length
            *    14263    14362: gap of unknown length
            *    14363    15984: contig of 1622 bp in length
            *    15985    16084: gap of unknown length
            *    16085    17502: contig of 1418 bp in length
            *    17503    17602: gap of unknown length
            *    17603    18902: contig of 1300 bp in length
```

```
            *    18903    19002: gap of unknown length
            *    19003    20008: contig of 1006 bp in length
            *    20009    20108: gap of unknown length
            *    20109    21372: contig of 1264 bp in length
            *    21373    21472: gap of unknown length
            *    21473    22675: contig of 1203 bp in length
            *    22676    22775: gap of unknown length
            *    22776    24088: contig of 1313 bp in length
            *    24089    24188: gap of unknown length
            *    24189    25354: contig of 1166 bp in length
            *    25355    25454: gap of unknown length
            *    25455    26641: contig of 1187 bp in length
            *    26642    26741: gap of unknown length
            *    26742    29340: contig of 2599 bp in length
            *    29341    29440: gap of unknown length
            *    29441    31609: contig of 2169 bp in length
            *    31610    31709: gap of unknown length
            *    31710    33695: contig of 1986 bp in length
            *    33696    33795: gap of unknown length
            *    33796    35117: contig of 1322 bp in length
            *    35118    35217: gap of unknown length
            *    35218    38766: contig of 3549 bp in length
            *    38767    38866: gap of unknown length
            *    38867    43795: contig of 4929 bp in length
            *    43796    43895: gap of unknown length
            *    43896    45218: contig of 1323 bp in length
            *    45219    45318: gap of unknown length
            *    45319    50886: contig of 5568 bp in length
            *    50887    50986: gap of unknown length
            *    50987    57066: contig of 6080 bp in length
            *    57067    57166: gap of unknown length
            *    57167    64436: contig of 7270 bp in length
            *    64437    64536: gap of unknown length
            *    64537    74858: contig of 10322 bp in length
            *    74859    74958: gap of unknown length
            *    74959    86302: contig of 11344 bp in length
            *    86303    86402: gap of unknown length
            *    86403    96560: contig of 10158 bp in length
            *    96561    96660: gap of unknown length
            *    96661   130540: contig of 33880 bp in length.
FEATURES             Location/Qualifiers
     source          1. .130540
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-102A24"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    37619 a  23855 c  30292 g  28886 t   9888 others
ORIGIN


  Query Match             3.3%;  Score 76.2;  DB 76;  Length 130540;
  Best Local Similarity   48.1%;  Pred. No. 0.0047;
  Matches  245;  Conservative    0;  Mismatches  263;  Indels    1;  Gaps    1;

Qy    215 ttacttttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaatagg 274
          ||  ||||||     |  || | ||||   | ||| |    | ||        |||||
Db  23437 TTTATTTTTGATTTTATATATTAATTATTTAGTTTTTACTTTTTGAATTTTTAAAATATA 23378

Qy    275 ttctttatttttggcatgtatgccttttttatttttcattcaatttttttttttttctcagac 334
          ||  |||| |||  ||| |||   ||||||| |    |  | ||||| |||| | | |
Db  23377 TTTATTATCTTTATCAT-TATATTTTTTTATATATTTCTATTTTTTTATTTTATATTATTG 23319

Qy    335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
          |       || |       |    |    |  |  || |   | |   |    | ||  |
Db  23318 ATTGTGTTTAATTTAATTTTATTTAATTATTTTGATGTTTATGTATATTATTTGTTATTT 23259

Qy    395 gtcagtctactttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
            |  | |  ||||        | | |||  ||    ||       |  || || |
Db  23258 ATATTTATTTTTTTATTATCTTTTTCTTTTTATTTATTTTTTTTTTTTTATGTTTTATTT 23199

Qy    455 tacttctttattattgataatttttattttcatttttttgctttcattattatacatatttt 514
          || || ||| || ||  |  | || | || || ||||  ||  ||| || |   ||| ||
```

```
Db  23198 TAATTTTTTTTTTTTTTTTTTATTGTGTTAATATTTTTTTTATATTTTTTTTTTTTATGTT 23139

Qy    515 ggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacact 574
            |  |       || |  ||| |   |       |  | ||||||||     ||   | |
Db  23138 TTTTTATTTTTTATTTGATTTTATATTATTTTTTAATTATTTTATTATTTATTTATTTATT 23079

Qy    575 cctttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
            ||||       | |        ||| ||  || | |   |  | | | | ||  | ||| |||
Db  23078 TTTTTTACTTTTTTTATTTTTTTTTTTTTTGATTTTTTTATTTTTTTGATTTATTTTTTTT 23019

Qy    635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
          ||||  |||| |  || | | |  |||||||||  ||    || |  |||| |      | |
Db  23018 TTTTTATTTTATATATTTAATTTATTTTATTTTTTTTTTTTATTTTTTAGTTTTTTTTATG 22959

Qy    695 ggtacggaacttttttttaatgaggaattt 723
            |     | ||||||||    ||    |||
Db  22958 TATTTTTTAGTTTTTTTTATGTAGTTTTTT 22930


RESULT  15
AC024285
LOCUS       AC024285    83440 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-22F23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC024285
VERSION     AC024285.1  GI:7107995
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 83440)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-22F23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 83440)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4112
                Center clone name: 22_F_23
            --------------
            * NOTE: This record contains 86 individual
            * sequencing reads that have not been assembled into
```

```
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      868: contig of 868 bp in length
*      869 968: gap of      100 bp
*      969     1849: contig of 881 bp in length
*     1850 1949: gap of      100 bp
*     1950     2798: contig of 849 bp in length
*     2799 2898: gap of      100 bp
*     2899     3794: contig of 896 bp in length
*     3795 3894: gap of      100 bp
*     3895     4785: contig of 891 bp in length
*     4786 4885: gap of      100 bp
*     4886     5764: contig of 879 bp in length
*     5765 5864: gap of      100 bp
*     5865     6712: contig of 848 bp in length
*     6713 6812: gap of      100 bp
*     6813     7702: contig of 890 bp in length
*     7703 7802: gap of      100 bp
*     7803     8701: contig of 899 bp in length
*     8702 8801: gap of      100 bp
*     8802     9666: contig of 865 bp in length
*     9667 9766: gap of      100 bp
*     9767    10628: contig of 862 bp in length
*    10629 10728: gap of      100 bp
*    10729    11589: contig of 861 bp in length
*    11590 11689: gap of      100 bp
*    11690    12556: contig of 867 bp in length
*    12557 12656: gap of      100 bp
*    12657    13530: contig of 874 bp in length
*    13531 13630: gap of      100 bp
*    13631    14504: contig of 874 bp in length
*    14505 14604: gap of      100 bp
*    14605    15484: contig of 880 bp in length
*    15485 15584: gap of      100 bp
*    15585    16457: contig of 873 bp in length
*    16458 16557: gap of      100 bp
*    16558    17434: contig of 877 bp in length
*    17435 17534: gap of      100 bp
*    17535    18418: contig of 884 bp in length
*    18419 18518: gap of      100 bp
*    18519    19392: contig of 874 bp in length
*    19393 19492: gap of      100 bp
*    19493    20309: contig of 817 bp in length
*    20310 20409: gap of      100 bp
*    20410    21298: contig of 889 bp in length
*    21299 21398: gap of      100 bp
*    21399    22260: contig of 862 bp in length
*    22261 22360: gap of      100 bp
*    22361    23214: contig of 854 bp in length
*    23215 23314: gap of      100 bp
*    23315    24174: contig of 860 bp in length
*    24175 24274: gap of      100 bp
*    24275    25129: contig of 855 bp in length
*    25130 25229: gap of      100 bp
*    25230    26098: contig of 869 bp in length
*    26099 26198: gap of      100 bp
*    26199    27086: contig of 888 bp in length
*    27087 27186: gap of      100 bp
*    27187    28080: contig of 894 bp in length
*    28081 28180: gap of      100 bp
*    28181    29047: contig of 867 bp in length
*    29048 29147: gap of      100 bp
*    29148    30030: contig of 883 bp in length
*    30031 30130: gap of      100 bp
*    30131    30997: contig of 867 bp in length
*    30998 31097: gap of      100 bp
*    31098    31943: contig of 846 bp in length
*    31944 32043: gap of      100 bp
*    32044    32891: contig of 848 bp in length
*    32892 32991: gap of      100 bp
*    32992    33863: contig of 872 bp in length
*    33864 33963: gap of      100 bp
*    33964    34818: contig of 855 bp in length
*    34819 34918: gap of      100 bp
*    34919    35762: contig of 844 bp in length
*    35763 35862: gap of      100 bp
*    35863    36727: contig of 865 bp in length
*    36728 36827: gap of      100 bp
*    36828    37697: contig of 870 bp in length
*    37698 37797: gap of      100 bp
*    37798    38667: contig of 870 bp in length
*    38668 38767: gap of      100 bp
*    38768    39634: contig of 867 bp in length
*    39635 39734: gap of      100 bp
*    39735    40586: contig of 852 bp in length
*    40587 40686: gap of      100 bp
*    40687    41577: contig of 891 bp in length
*    41578 41677: gap of      100 bp
*    41678    42552: contig of 875 bp in length
*    42553 42652: gap of      100 bp
*    42653    43517: contig of 865 bp in length
*    43518 43617: gap of      100 bp
*    43618    44482: contig of 865 bp in length
*    44483 44582: gap of      100 bp
*    44583    45460: contig of 878 bp in length
*    45461 45560: gap of      100 bp
*    45561    46416: contig of 856 bp in length
*    46417 46516: gap of      100 bp
*    46517    47415: contig of 899 bp in length
*    47416 47515: gap of      100 bp
*    47516    48421: contig of 906 bp in length
*    48422 48521: gap of      100 bp
*    48522    49395: contig of 874 bp in length
*    49396 49495: gap of      100 bp
*    49496    50337: contig of 842 bp in length
*    50338 50437: gap of      100 bp
*    50438    51311: contig of 874 bp in length
*    51312 51411: gap of      100 bp
*    51412    52281: contig of 870 bp in length
*    52282 52381: gap of      100 bp
*    52382    53266: contig of 885 bp in length
*    53267 53366: gap of      100 bp
*    53367    54233: contig of 867 bp in length
*    54234 54333: gap of      100 bp
*    54334    55219: contig of 886 bp in length
*    55220 55319: gap of      100 bp
*    55320    56191: contig of 872 bp in length
*    56192 56291: gap of      100 bp
*    56292    57136: contig of 845 bp in length
*    57137 57236: gap of      100 bp
*    57237    58132: contig of 896 bp in length
*    58133 58232: gap of      100 bp
*    58233    59105: contig of 873 bp in length
*    59106 59205: gap of      100 bp
*    59206    60078: contig of 873 bp in length
*    60079 60178: gap of      100 bp
*    60179    61042: contig of 864 bp in length
*    61043 61142: gap of      100 bp
*    61143    62035: contig of 893 bp in length
*    62036 62135: gap of      100 bp
*    62136    63021: contig of 886 bp in length
*    63022 63121: gap of      100 bp
*    63122    64006: contig of 885 bp in length
*    64007 64106: gap of      100 bp
*    64107    64974: contig of 868 bp in length
*    64975 65074: gap of      100 bp
*    65075    65944: contig of 870 bp in length
*    65945 66044: gap of      100 bp
*    66045    66922: contig of 878 bp in length
```

```
          *    66923 67022: gap of      100 bp
          *    67023    67919: contig of 897 bp in length

  Query Match             3.3%;  Score 75.8;  DB 69;  Length 83440;
  Best Local Similarity   42.6%;  Pred. No. 0.0057;
  Matches  203;  Conservative    0;  Mismatches  273;  Indels    0;  Gaps    0;

Qy     275 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 334
           || | | |||||    | | |   ||| | ||||  ||     || |||||||  |
Db   45900 TTNTNTTTTTTTTNTTTTTTTTTTTNTTTNTTTTTTTNTTTTTTTNNTTNTTTTTTTTTTTTTN 45959

Qy     335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
             |  |  |       |    ||       |   | |   | || ||    | ||| |
Db   45960 TTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTNTTTTTTTTTNNTTTTTTTTNTNTTTTTTTT 46019

Qy     395 gtcagtctacttttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
             | |  |||||  |    | || ||   |    ||       |   | || |
Db   46020 TNTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTNTTT 46079

Qy     455 tacttctttattattgataattttattttcatttttgctttcattattatacatatttt 514
           |  || ||| || ||  |  |||| ||||  ||  ||  |||  || || |   | | ||
Db   46080 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTNTT 46139

Qy     515 ggtggagaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacact 574
             |       ||   |||| |              ||| || |     |       |
Db   46140 TTTNTNTNTTTTTTTNNTTTTTTNTNTTTTTTTNNNNTTTTTTTTTTTTTTTTTTTTTTTT 46199

Qy     575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
             ||||||    | ||      ||| ||   | | |       | |       ||| |||
Db   46200 TTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTNTTTTTTTNTNNTTTTNNNNTTTTTTTTTT 46259

Qy     635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
           ||||  |||    | | |   || ||   | ||   || |  ||||        | || |
Db   46260 TTTTTTTTTTTTTNTNTTTTTTTTTTTTNNNTTTTTTTTTTTTNNTTTTNTTNTGTTTNTTTT 46319

Qy     695 ggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttctctctt 750
             |       ||||||  |      |||  |  |  ||| ||  |||| | | ||
Db   46320 TTTNNNTNNTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTGGGTTTTTTTTTTTTT 46375


Search completed: April 30, 2001, 19:08:52
Job time: 37006 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:04:50 ; Search time 933.78 Seconds
(without alignments)
1454.796 Million cell updates/sec

Title: US-09-721-183-3
Perfect score: 2327
Sequence: 1 gatggatgcatctcaaaatg.....actcaattctggttttcccc 2327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2257.6	97.0	10319	21	C74759	Human ORFX ORF314
2		2257.6	97.0	10320	21	A54123	Breast cancer prot
3		992.6	42.7	1009	21	C79479	cDNA sequence for
4		305	13.1	333	19	V27248	HeLa gene sequence
5		240	10.3	674	22	F44875	Human breast cance
6		169.4	7.3	173	21	C21536	Human secreted pro
c	7	70.8	3.0	19124	18	T72882	Plasmodium var-7 g
c	8	70.8	3.0	19124	21	Z98287	Plasmodium var-7 p
c	9	69	3.0	6644	20	X33181	Base sequence of t
c	10	69	3.0	7372	20	X33182	Base sequence of t
c	11	69	3.0	7797	20	X33180	Cowpox virus bsr f
c	12	69	3.0	7996	20	X33184	Base sequence of t
c	13	65.6	2.8	9789	17	T41852	cDNA encoding Plas
c	14	64.6	2.8	2364	21	A70246	Plasmodium falcipa
	15	63	2.7	76	21	C28768	Human secreted pro
	16	61.2	2.6	58407	19	V21210	Methanococcus jann
c	17	60.8	2.6	5454	21	A70236	Plasmodium falcipa
c	18	60.2	2.6	11922	21	A70187	Plasmodium falcipa
c	19	58.8	2.5	875	21	A01920	Human colon cancer
	20	58.6	2.5	5852	12	Q11710	Dictyostelium plas
c	21	57.6	2.5	1251	21	A70119	Plasmodium falcipa
c	22	57.4	2.5	2503	15	Q53480	pNPX30 xylanase cD
	23	57.4	2.5	19124	18	T72882	Plasmodium var-7 g
	24	57.4	2.5	19124	21	Z98287	Plasmodium var-7 p
c	25	57	2.4	4590	7	N60472	Sequence encoding
c	26	54.2	2.3	6033	21	A70152	Plasmodium falcipa
c	27	53.6	2.3	1035	21	A01936	Human colon cancer
	28	52.6	2.3	4590	7	N60472	Sequence encoding
c	29	52.6	2.3	6124	11	Q03568	Sequence encoding
c	30	52.4	2.3	1612	7	N60392	Sequence encoding
c	31	52	2.2	6243	20	Z09598	Clostridium specie
c	32	51.8	2.2	3567	21	A70117	Plasmodium falcipa
c	33	51.8	2.2	5940	21	A70105	Plasmodium falcipa
c	34	51.4	2.2	7620	21	A70132	Plasmodium falcipa
c	35	51.2	2.2	58407	19	V21210	Methanococcus jann
c	36	51	2.2	909	21	A01930	Human colon cancer
	37	51	2.2	2104	13	Q25273	Sequence encoding
	38	51	2.2	2486	21	A97037	Nucleotide sequenc
c	39	50.8	2.2	1267	20	Z17188	Human gene express
c	40	50.8	2.2	3582	21	A70241	Plasmodium falcipa
c	41	50.8	2.2	6042	21	A70199	Plasmodium falcipa
c	42	50.4	2.2	1199	20	X08695	Novel nucleotide s
c	43	50.4	2.2	1711	19	V33136	Plasmodium berghei
	44	50.4	2.2	26811	20	X20253	Borrelia burgdorfe
c	45	50	2.1	7503	21	A70206	Plasmodium falcipa

ALIGNMENTS

RESULT 1
C74759
ID C74759 standard; cDNA; 10319 BP.
XX
AC C74759;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF314 polynucleotide sequence SEQ ID NO:627.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; B40550.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 743-750; 5507pp; English.
XX
CC C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 10319 BP; 3176 A; 2006 C; 2093 G; 3043 T; 1 other;

Query Match 97.0%; Score 2257.6; DB 21; Length 10319;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2293; Conservative 0; Mismatches 32; Indels 3; Gaps 3;

```
Qy      1 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6905 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 6964

Qy     61 agagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6965 agagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaat 7024

Qy    121 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7025 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 7084

Qy    181 agaagtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactcatt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7085 agaagtttggttaacttgggcaatatgacaagtattacttttttgggcaaaactactcatt 7144

Qy    241 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgcctt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7145 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgcctt 7204

Qy    301 tttattttcattcaatttttttttttttctcagacagacatagtagtaacgactagcattg 360
          |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db   7205 tttattttcattcaatttttttttttttctcagacagacatagtagtatcaactagcattg 7264

Qy    361 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtagaatatt 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   7265 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaatatt 7324

Qy    421 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7325 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 7384

Qy    481 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7385 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 7444

Qy    541 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7445 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 7504

Qy    601 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7505 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 7564

Qy    661 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7565 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 7624

Qy    721 tttcatgatgatttaggaatttttctctcttggaaaaggcttcccctgtgatgaaaatgat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7625 tttcatgatgatttaggaatttttctctcttggaaaaggcttcccctgtgatgaaaatgat 7684

Qy    781 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7685 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 7744

Qy    841 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7745 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 7804

Qy    901 tccttttctcactctttctctcancactcacctctgtattcattctgttgtttgggatag 960
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   7805 tccttttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatag 7864

Qy    961 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7865 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 7924

Qy   1021 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7925 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 7984

Qy   1081 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7985 agagtatgcccaattctctctacatgacatattgagatttttttaatcaacttttaaga 8044

Qy   1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8045 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 8104

Qy   1201 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8105 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 8164

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8165 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 8224

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgcctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8225 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatgcctg 8284

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8285 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 8344

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8345 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 8404

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccag 1560
```

```
Db   8405 gctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccag 8464

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8465 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 8524

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcnnn 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8525 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcaaa 8584

Qy   1681 nnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagcttgct 1740
                                  ||||||||||||||||||||||||||||||||||||
Db   8585 aaaaaaaaaaaaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgct 8644

Qy   1741 gactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaa 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8645 gactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaa 8704

Qy   1801 atattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8705 atattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatg 8764

Qy   1861 agaagaagatagaaggacgcccgtcgggactttggagggcagtgttatttttcccaaagaa 1920
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   8765 agaagaagatagaaggacgcccgtcgggactttggagggcattgttatttttcccaaagaa 8824

Qy   1921 agacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagta 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8825 agacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagta 8884

Qy   1981 ctgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8885 ctgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcat 8944

Qy   2041 cccagtt-attgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagta 2099
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8945 cccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagta 9004

Qy   2100 tgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgt 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9005 tgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgt 9064

Qy   2160 acattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacacc 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9065 acattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacacc 9124

Qy   2220 ggagaaaagcaggaataactttaccgtggaaataatgtttagcttttatgcagagaaaat 2279
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   9125 ggagaaaagcaggaataactttaccgtggaaataatgtttagcttttat-cagagaaaat 9183

Qy   2280 tgtccttcctagagcatagagtcccaaaactcaattctggttttcccc 2327
          ||||||| ||||||||||||||||||||||||||||||||||||||
Db   9184 tgtcctt-ctagagcatagagtcccaaaactcaattctggttttcccc 9230
```

```
RESULT   2
A54123
ID   A54123 standard; DNA; 10320 BP.
XX
AC   A54123;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Breast cancer protein BCN2 coding sequence.
XX
KW   Breast cancer; diagnosis; prognosis; detection; screening;
KW   antibody; oestrogen receptor; anti-oestrogen; immune response;
KW   lymph node; metastases; tumour; BCR3; BCQ8; BCQ5; BCH1; BCN1; BCN2;
KW   BCN5; BCO2; BCX2; BCX3; BCA2; BCR2; BCJ7; BCY3; human; ds.
XX
OS   Homo sapiens.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             645..4490
FT                   /*tag= a
FT                   /product= BCN2 protein
XX
PN   WO200055629-A2.
XX
PD   21-SEP-2000.
XX
PF   15-MAR-2000; 2000WO-US06952.
XX
PR   15-MAR-1999;   99US-0268865.
PR   12-NOV-1999;   99US-0439878.
PR   12-NOV-1999;   99US-0440370.
PR   15-NOV-1999;   99US-0440493.
PR   16-NOV-1999;   99US-0440676.
PR   16-NOV-1999;   99US-0440677.
PR   29-NOV-1999;   99US-0450810.
PR   02-DEC-1999;   99US-0453137.
PR   08-MAR-2000; 2000US-0453137.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack D,  Gish KC;
XX
DR   WPI; 2000-638216/61.
DR   P-PSDB; B00190.
XX
PT   Screening drug candidates for their ability to modulate breast cancer
PT   by contacting the drug to a cell expressing an expression profile gene
PT   and determining modulation of expression of the gene
XX
PS   Disclosure; Fig 46A-D; 258pp; English.
XX
CC   New methods for screening drug candidates are described which
CC   comprise adding a drug candidate to a cell that expresses a protein
CC   selected from BCH1, BCA2, BCJ7, BCN1, BCN5, BCO2, BCQ5, BCR2, BCX2
CC   and BCY3 or their fragments and determining the effect of the drug
CC   on the expression of those proteins. Antibodies to breast cancer
CC   genes (specifically BCH1 or its fragment (BCH1p1 or BCH1p2)) are
CC   useful for inhibiting and treating breast cancer in individuals who
CC   are non-responsive to anti-oestrogen and positive for oestrogen
CC   receptor. Compositions comprising BCH1 or a nucleic acid encoding
CC   BCH1 are useful for eliciting an immune response in an individual.
CC   The antibodies are also useful for the diagnosis and prognosis of
CC   breast cancer and for screening compositions which modulate the
CC   breast cancer phenotype. The method allows rapid and simple
CC   detection of lymph node metastases.
XX
SQ   Sequence 10320 BP; 3174 A; 2005 C; 2096 G; 3042 T; 3 other;
```

```
  Query Match             97.0%;  Score 2257.6;  DB 21;  Length 10320;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 2293;  Conservative    0;  Mismatches   32;  Indels    3;  Gaps    3;

Qy       1 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6904 gatggatgcatctcaaaatgtatagccagacttgagaggtgacaattaaagatctaaaaa 6963

Qy      61 agagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaat 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6964 agagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataacagaat 7023

Qy     121 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7024 gcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaaaacacc 7083

Qy     181 agaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactcatt 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7084 agaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaactactcatt 7143

Qy     241 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgcctt 300
```

```
Db   7144 aagcaatttctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttt 7203

Qy    301 tttattttcattcaatttttttttttttctcagacagacatagtagtaacgactagcattg 360
          ||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db   7204 tttattttcattcaatttttttttttttctcagacagacatagtagtatcaactagcattg 7263

Qy    361 gaaaatacatatcactattcttggaatatttatggtcagtctactttttagtagaatatt 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   7264 gaaaatacatatcactattcttggaatatttatggtcagtctacttttttagtaaaatatt 7323

Qy    421 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7324 tttggatagcgttgacacgatagatcttattccatacttctttattattgataatttttat 7383

Qy    481 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7384 tttcatttttttgctttcattattatacatattttggtggagaagaggttgggctttttttg 7443

Qy    541 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7444 aaagagacaaaaatttattataacactaaacactccttttttgacatattaaagccttta 7503

Qy    601 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7504 ttccatctctcaagatatattataaaatttatttttttaatttaagatttctgaattatt 7563

Qy    661 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7564 ttatcttaaattgtgattttaaacgagctattatggtacggaacttttttttaatgaggaa 7623

Qy    721 tttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgat 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7624 tttcatgatgatttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgat 7683

Qy    781 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7684 gtgccagctaaaattgtgtgccatttaaaaactgaaaatattttaaaattatttgtctat 7743

Qy    841 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7744 attctaaattgagctttggatcaaactttaggccaggaccagctcatgcgttctcattct 7803

Qy    901 tccttttctcactctttctctcancactcacctctgtattcattctgttgtttgggatag 960
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   7804 tccttttctcactctttctctcatcactcacctctgtattcattctgttgtttgggatag 7863

Qy    961 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7864 aaaaatcataaagagccaacccatctcagaacgttgtggattgagagagacactacatga 7923

Qy   1021 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7924 ctccaagtatatgagaaaaggacagagctctaattgataactctgtagttcaaaaggaaa 7983

Qy   1081 agagtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaaga 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7984 agagtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaaga 8043

Qy   1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8044 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 8103

Qy   1201 atggggattctttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8104 atggggattctttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 8163

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8164 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatt 8223

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8224 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctg 8283

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8284 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 8343

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8344 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 8403

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8404 gctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatccag 8463

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8464 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 8523

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcnnn 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8524 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggcaaa 8583

Qy   1681 nnnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagcttgct 1740
                                    ||||||||||||||||||||||||||||||||||||
Db   8584 aaaaaaaaaagaaaaaaagatcaagtcgtctttggtgagccagtaaggtgaaagcttgct 8643

Qy   1741 gactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaa 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8644 gactgtccaaggcacaagagaaaattgaggaattgaaatgcaacctgagtatcaaactaa 8703

Qy   1801 atattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8704 atattctaatcaaaggtaggtactgttaggtggaattctatcagcaggcaactgcaaatg 8763

Qy   1861 agaagaagatagaaggacgcccgtcgggactttggagggcagtgttatttcccaaagaa 1920
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   8764 agaagaagatagaaggacgcccgtcgggactttggagggcattgttatttcccaaagaa 8823

Qy   1921 agacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagta 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8824 agacggccaagggcagaggcatggattctttgcagagcacttccttttggtttttcagta 8883

Qy   1981 ctgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcat 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8884 ctgtttcatagacagtgggctcacatgttcctgatagtgctgcagttgcttagaaagcat 8943

Qy   2041 cccagtt-attgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagta 2099
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8944 cccagttaattgcagtaattagaacttctggaatatgctagggcagaagtatgtcaagta 9003

Qy   2100 tgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgt 2159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9004 tgtcacatgaagaaaatgtgaaattcaagagtaatccacacgtgagaaactagacaatgt 9063

Qy   2160 acattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacacc 2219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9064 acattcatgtgttctcttgaaaggaaagggagagctgtaagcttcactctgtcctacacc 9123

Qy   2220 ggagaaaagcaggaataactttaccgtggaaataatgtttagctttttatgcagagaaaat 2279
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   9124 ggagaaaagcaggaataactttaccgtggaaataatgtttagctttttat-cagagaaaat 9182

Qy   2280 tgtccttcctagagcatagagtcccaaaactcaattctggtttttcccc 2327
          ||||||| ||||||||||||||||||||||||||||||||||||||||
Db   9183 tgtcctt-ctagagcatagagtcccaaaactcaattctggtttttcccc 9229
```

```
RESULT   3
C79479
ID   C79479 standard; cDNA; 1009 BP.
XX
AC   C79479;
XX
```

```
DT   07-FEB-2001  (first entry)
XX
DE   cDNA sequence for human breast tumour clone #19606.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy;
KW   breast cancer; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09688.
XX
PR   09-APR-1999;   99US-0288950.
PR   02-JUL-1999;   99US-0346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer -
XX
PS   Claim 4; Page 95; 95pp; English.
XX
CC   The present sequence was isolated from a breast tumour cDNA library. It
CC   is provided in a specification relating to compounds for immunotherapy
CC   and diagnosis of breast cancer. Breast tumour antigens and the
CC   polynucleotides that encode them may be used in the production of a
CC   pharmaceutical composition to be used in the treatment of breast cancer.
CC   Proliferated T cells and incubated antigen presenting cells are also
CC   required. The polypeptides and polynucleotides may also be used to
CC   produce a vaccine.
XX
SQ   Sequence 1009 BP; 316 A; 163 C; 193 G; 336 T; 1 other;
```

```
  Query Match          42.7%;  Score 992.6;  DB 21;  Length 1009;
  Best Local Similarity  99.8%;  Pred. No. 8.1e-187;
  Matches  992;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    684 cgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatttt 743
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatttt 60

Qy    744 ctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgcca 803
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgcca 120

Qy    804 tttaaaaactgaaaatattttaaaattatttgtctatattctaaattgagctttggatca 863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tttaaaaactgaaaatattttaaaattatttgtctatattctaaattgagctttggatca 180

Qy    864 aactttaggccaggaccagctcatgcgttctcattcttccttttctcactctttctctca 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aactttaggccaggaccagctcatgcgttctcattcttccttttctcactctttctctca 240

Qy    924 ncactcacctctgtattcattctgttgtttgggatagaaaaatcataaagagccaaccca 983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcactcacctctgtattcattctgttgtttgggatagaaaaatcataaagagccaaccca 300

Qy    984 tctcagaacgttgtggattgagagagacactacatgactccaagtatatgagaaaaggac 1043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 tctcagaacgttgtggattgagagagacactacatgactccaagtatatgagaaaaggac 360

Qy   1044 agagctctaattgataactctgtagttcaaaaggaaaagagtatgcccaattctctctac 1103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 agagctctaattgataactctgtagttcaaaaggaaaagagtatgcccaattctctctac 420

Qy   1104 atgacatattgagatttttttttaatcaacttttaagatagtgatgttctgttctaaaactg 1163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 atgacatattgagatttttttttaatcaacttttaagatagtgatgttctgttctaaaactg 480

Qy   1164 ttctgttttagtgaaggtagatttttataaaaacaagcatggggattctttttctaaggtaa 1223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttctgttttagtgaaggtagatttttataaaaacaagcatggggattctttttctaaggtaa 540

Qy   1224 tattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcac 1283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db    541 tattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcmc 600

Qy   1284 attatgtttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctc 1343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 attatgtttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctc 660

Qy   1344 caaaaatattaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataa 1403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 caaaaatattaaatgtatatatattttaaaatgcctgaggaaatacatttttcttaataa 720

Qy   1404 actgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaa 1463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 actgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaa 780

Qy   1464 aacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaa 1523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 aacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaa 840

Qy   1524 ttaatgcatttctagaggggaatatctgccatccagtggtggaaatgtggagtaaagaa 1583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttaatgcatttctagaggggaatatctgccatccagtggtggaaatgtggagtaaagaa 900

Qy   1584 gctggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaa 1643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 gctggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaa 960

Qy   1644 ctttagctactgtctttggtttgagagccatggc 1677
          |||||||||||||||||||||||||||||||||
Db    961 ctttagctactgtctttggtttgagagccatggc 994
```

```
RESULT   4
V27248
ID   V27248 standard; DNA; 333 BP.
XX
AC   V27248;
XX
DT   28-SEP-1998  (first entry)
XX
DE   HeLa gene sequence E7_10_9.
XX
KW   Gene identification; HeLa; Fas antibody; cancer; antisense;
KW   therapy; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO9821366-A1.
XX
PD   22-MAY-1998.
XX
PF   12-NOV-1997;   97WO-US20989.
XX
PR   13-NOV-1996;   96US-0030549.
XX
PA   (QBIQ-) QBI ENTERPRISES LTD.
XX
PI   Deiss LP,  Efimova E,  Einat P,  Vasquez-Iaslop NC;
PI   Yehiely F;
XX
DR   WPI; 1998-297966/26.
XX
PT   Identifying genes essential for maintenance of specific cell
```

PT phenotypes - using gene inactivation, positive selection and
PT subtraction analysis, useful to develop pharmaceuticals to undesired
PT phenotypes e.g. cancer
XX
PS Example 2a; Page 51; 66pp; English.
XX
CC E7_10_9 is one of 22 gene sequences (see V27227-48) identified
CC in HeLa cells as being involved in Fas antibody sensitivity. A
CC novel method of the invention was applied to HeLa cells treated
CC with anti-Fas antibody in order to identify genes that when
CC knocked-out cause sensitisation of HeLa cells to the action of
CC anti-Fas antibodies. E7_10_9 shows no good match against known
CC genes in a combined nonredundant database, but does show good
CC matches against 2 ESTs (gb R54192 and gb H39863). Sequence
CC analysis indicates that the fragment inserted in the expression
CC library was in the antisense orientation. Thus, reduction of
CC the gene product that is encoded by the full-length message
CC representing this clone leads to supersensitisation of cells to
CC treatment with anti-Fas antibodies.
XX
SQ Sequence 333 BP; 100 A; 58 C; 83 G; 92 T; 0 other;

Query Match 13.1%; Score 305; DB 19; Length 333;
Best Local Similarity 99.7%; Pred. No. 1.6e-51;
Matches 316; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy   1887 ggactttggagggcagtgttatttttcccaaagaaagacggccaagggcagaggcatggat 1946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 ggactttggagggcagtgttatttttcccaaagaaagacggccaagggcagaggcatggat 64

Qy   1947 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 2006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 124

Qy   2007 gttcctgatagtgctgcagttgcttagaaagcatcccagtt-attgcagtaattagaact 2065
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    125 gttcctgatagtgctgcagttgcttagaaagcatcccagttaattgcagtaattagaact 184

Qy   2066 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 2125
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 244

Qy   2126 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 2185
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 304

Qy   2186 agggagagctgtaagct 2202
          |||||||||||||||||
Db    305 agggagagctgtaagct 321
```

RESULT 5
F44875
ID F44875 standard; cDNA; 674 BP.
XX
AC F44875;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human breast cancer related protein coding sequence SEQ ID NO: 31.
XX
KW Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17536.
XX
PR 23-JUN-1999; 99US-0140903.
PR 12-OCT-1999; 99US-0158980.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Mitcham JL;
XX
DR WPI; 2001-041426/05.
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
PT prevention, treatment and diagnosis of breast cancer -
XX
PS Claim 5; Page 128; 165pp; English.
XX
CC The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX
SQ Sequence 674 BP; 222 A; 110 C; 128 G; 213 T; 1 other;

Query Match 10.3%; Score 240; DB 22; Length 674;
Best Local Similarity 99.2%; Pred. No. 1.2e-38;
Matches 262; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```
Qy   2064 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 2123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cttctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaat 60

Qy   2124 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 2183
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaagg 120

Qy   2184 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 2243
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaagggagagctgtaagcttcactctgtcctacaccggagaaaagcaggaataactttac 180

Qy   2244 cgtggaaataatgtttagcttttatgcagagaaaattgtccttcctagagcatagagtcc 2303
          |||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db    181 cgtggaaataatgtttagcttttat-cagagaaaattgtcctt-ctagagcatagagtcc 238

Qy   2304 caaaactcaattctggttttcccc 2327
          |||||||||||||||||||||||
Db    239 caaaactcaattctggttttcccc 262
```

RESULT 6
C21536
ID C21536 standard; cDNA; 173 BP.
XX
AC C21536;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 25611.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX

```
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 25611; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 173 BP; 58 A; 25 C; 30 G; 55 T; 5 other;


  Query Match             7.3%;  Score 169.4;  DB 21;  Length 173;
  Best Local Similarity   96.5%;  Pred. No. 7.2e-25;
  Matches  167;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1174 gtgaaggtagatttttataaaacaagcatggggattcttttctaaggtaatattaatgag 1233
           ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1 gtgaaggtagattttatamaacaagcatggggattcttttctaaggtaatattaatgag 60

Qy    1234 aagggaaaaaagtatctttaacagctctttgttgaagcctgtggtagcacattatgttta 1293
           ||||||||||||||||||||||||||||||:||||||||||:||||||||||||||||||
Db      61 aagggaaaaaagtatctttaacagctctttrttgaagcctgtrgtagcacattatgttta 120

Qy    1294 taattgcacatgtgcacataatctattatgatccaatgcaaatacagctccaa 1346
           |||||:|||||||||||||||||||||||||||||||||||||||||||| |||:
Db     121 taattrcacatgtgcacataatctattatgatccaatgcaaatacagccccar 173


RESULT   7
T72882/c
ID   T72882 standard; cDNA; 19124 BP.
XX
AC   T72882;
XX
DT   12-SEP-1997  (first entry)
XX
DE   Plasmodium var-7 gene.
XX
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
XX
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   exon            7317..15139
FT                   /*tag= a
FT                   /number= 1
FT   intron          15140..16205
FT                   /*tag= b
FT                   /number= 1
FT   exon            16206..17552
FT                   /*tag= c
FT                   /number= 2
FT                   /note= "no stop codon given"
XX
PN   WO9640766-A2.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09508.
XX
PR   07-JUN-1995;   95US-0487826.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Chitnis C,  Miller LH,  Peterson DS,  Sim KL,  Su X;
PI   Wellems TE;
XX
DR   WPI; 1997-052231/05.
DR   P-PSDB; W22475.
XX
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
XX
PS   Claim 4; Page 56-61; 96pp; English.
XX
CC   This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and
CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
XX
SQ   Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;


  Query Match             3.0%;  Score 70.8;  DB 18;  Length 19124;
  Best Local Similarity   48.9%;  Pred. No. 4.5e-05;
  Matches  256;  Conservative    0;  Mismatches  257;  Indels   11;  Gaps    2;

Qy     163 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattacttttt 222
           |||  |||  |||       || |||  || | ||    ||  |      | |    |||
Db   15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy     223 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 282
           |     |    | | | |     ||   || | | |   |   |||||  || ||| |
Db   15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy     283 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 342
           |||    | |||   |||||  ||  ||||   ||  ||||||||          |  |
Db   15800 ATTTTATGTATATATTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy     343 tagtaacgactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 402
           |  ||   | |   |||  || |||  | |  || || ||    | ||||||  |   | |
Db   15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy     403 acttttagtagaatatttttggatagcgttgacacgatagatcttattccatac---tt 459
             |||||  |  ||| |||||   |  | |        |  || || ||    |    ||
Db   15688 TTTTTTTTTTTTTAATGTTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTTTATAATCATTT 15629

Qy     460 ctttattattgataatttttattttcatttttgctttcattattatacatattttggtgg 519
           |||  ||||  | ||||||  ||||  |||||||  |   | | || | |||  ||||  |
Db   15628 TTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569
```

```
Qy     520 agaagaggttgggctttttgaaagagacaaaaatttattataacactaaacactccttt 579
Db   15568 ATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy     580 tttgacatattaaagcctttattccatctctcaagatatattataaaatttatttttta 639
Db   15508 TTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy     640 atttaagatttctgaattattttatcttaaattgtgattttaaa 683
Db   15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405


RESULT   8
Z98287/c
ID   Z98287 standard; DNA; 19124 BP.
XX
AC   Z98287;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Plasmodium var-7 polypeptide encoding DNA.
XX
KW   DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW   DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
KW   protozoacide; var-7; ds.
XX
OS   Plasmodium sp.
XX
PN   US5993827-A.
XX
PD   30-NOV-1999.
XX
PF   07-JUN-1995;   95US-0487826.
XX
PR   10-SEP-1993;   93US-0119677.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Sim KL,  Chitnis C,  Peterson DS,  Su X,  Wellems TE,  Miller LH;
XX
DR   WPI; 2000-194198/17.
DR   P-PSDB; Y77904.
XX
PT   Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT   falciparum erythrocyte binding proteins useful for vaccinating against
PT   malaria -
XX
PS   Disclosure; Columns 91-108; 93pp; English.
XX
CC   The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC   (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC   identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC   Binding Protein (SABP), which are soluble proteins that appear in the
CC   culture supernatant after erythrocytes infected with malaria release
CC   merozoites. Immunochemical studies indicate that DABP and SABP are the
CC   respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC   and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC   used to vaccinate against malaria, especially caused by P. falciparum.
CC   Immunization with the polypeptide provides effective protection against
CC   malaria. The present sequence represents the var-7 polypeptide encoding
CC   DNA.
XX
SQ   Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;


  Query Match             3.0%;  Score 70.8;  DB 21;  Length 19124;
  Best Local Similarity   48.9%;  Pred. No. 4.5e-05;
  Matches  256;  Conservative    0;  Mismatches  257;  Indels   11;  Gaps    2;

Qy     163 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattacttttt 222
Db   15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy     223 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 282
Db   15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy     283 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 342
Db   15800 ATTTTATGTATATATTTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy     343 tagtaacgactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 402
Db   15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy     403 acttttttagtagaatatttttggatagcgttgacacgatagatcttattccatac---tt 459
Db   15688 TTTTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTTGTTTTTATTTTTTTTTATAATCATTT 15629

Qy     460 ctttattattgataatttttattttcatttttttgctttcattattatacatattttggtgg 519
Db   15628 TTTTTTTATATAAAATTTTTTTTTTAATTTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569

Qy     520 agaagaggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttt 579
Db   15568 ATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy     580 tttgacatattaaagcctttattccatctctcaagatatattataaaatttatttttta 639
Db   15508 TTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy     640 atttaagatttctgaattattttatcttaaattgtgattttaaa 683
Db   15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405


RESULT   9
X33181/c
ID   X33181 standard; DNA; 6644 BP.
XX
AC   X33181;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-ires-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Cowpox virus.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP04010.
XX
PR   08-SEP-1997;   97JP-0259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell
XX
PS   Example 1; Page 38-41; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
```

```
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC   is used in an example from the present invention.
XX
SQ   Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;


  Query Match             3.0%;  Score 69;  DB 20;  Length 6644;
  Best Local Similarity   46.9%;  Pred. No. 8.3e-05;
  Matches  216;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy     275 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 334
           || ||| |||||    | | |   ||||| ||||  ||   ||||||||||| |
Db    4180 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4121

Qy     335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
           |  |  |      |    ||       |  | |   | || ||    | ||| |
Db    4120 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4061

Qy     395 gtcagtctactttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
           |   | |  |||||  |    | |||||   |    ||        |   | || ||
Db    4060 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4001

Qy     455 tacttctttattattgataattttattttcatttttgctttcattattatacatatttt 514
           |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |   | ||||
Db    4000 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3941

Qy     515 ggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaacact 574
           |          ||    ||||||                ||| || |     |     |
Db    3940 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3881

Qy     575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
             ||||||    | ||      ||| ||   | | |     | | || |    ||| |||
Db    3880 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3821

Qy     635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
           ||||  |||    ||| |   || |||| | ||   || |  ||||          | || |
Db    3820 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3761

Qy     695 ggtacggaacttttttaatgaggaatttcatgatgattta 735
           |       |||||||  |  | || | || || | ||||
Db    3760 TTTTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 3720


RESULT  10
X33182/c
ID   X33182 standard; DNA; 7372 BP.
XX
AC   X33182;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-Bcl-xl-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP04010.
XX
PR   08-SEP-1997;   97JP-0259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell
XX
PS   Example 2; Page 41-45; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and
CC   is used in an example from the present invention.
XX
SQ   Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 other;


  Query Match             3.0%;  Score 69;  DB 20;  Length 7372;
  Best Local Similarity   46.9%;  Pred. No. 8.5e-05;
  Matches  216;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy     275 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 334
           || ||| |||||    | | |   ||||| ||||  ||   ||||||||||| |
Db    4908 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4849

Qy     335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
           |  |  |      |    ||       |  | |   | || ||    | ||| |
Db    4848 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4789

Qy     395 gtcagtctactttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
           |   | |  |||||  |    | |||||   |    ||        |   | || ||
Db    4788 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4729

Qy     455 tacttctttattattgataattttattttcatttttgctttcattattatacatatttt 514
           |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |   | ||||
Db    4728 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4669

Qy     515 ggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaacact 574
           |          ||    ||||||                ||| || |     |     |
Db    4668 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4609

Qy     575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
             ||||||    | ||      ||| ||   | | |     | | || |    ||| |||
```

```
Db   4608 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4549

Qy    635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
          ||||  |||    ||| |   || |||| | ||   || |  ||||          | || |
Db   4548 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4489

Qy    695 ggtacggaacttttttaatgaggaatttcatgatgattta 735
          |       |||||||  |  | || | || || | ||||
Db   4488 TTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 4448


RESULT 11
X33180/c
ID   X33180 standard; DNA; 7797 BP.
XX
AC   X33180;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Cowpox virus bsr full length gene sequence.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Cowpox virus.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP04010.
XX
PR   08-SEP-1997;   97JP-0259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell
XX
PS   Example 1; Page 34-38; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the cowpox virus bsr gene which
CC   is used in an example from the present invention.
XX
SQ   Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;


 Query Match             3.0%;  Score 69;  DB 20;  Length 7797;
 Best Local Similarity   46.9%;  Pred. No. 8.6e-05;
 Matches  216;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy    275 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 334
          || ||| |||||    | | |   |||||  ||||  ||   ||||||||||||  |
Db   5333 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5274

Qy    335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
              |  |  |       |    ||       |   | |   | || ||    | ||| |
Db   5273 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5214

Qy    395 gtcagtctacttttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
          |   | |  |||||  |    | |||||   |    ||          |   | || ||
Db   5213 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5154

Qy    455 tacttctttattattgataatttttattttcatttttgctttcattattatacatatttt 514
          |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |   | ||||
Db   5153 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5094

Qy    515 ggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaacact 574
            |        ||    ||||||                |||  || |     |       |
Db   5093 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5034

Qy    575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
            ||||||    | ||      ||| ||   | | |     | | || |    ||| |||
Db   5033 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4974

Qy    635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
          ||||  |||    ||| |   || |||| | ||   || |  ||||          | || |
Db   4973 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4914

Qy    695 ggtacggaacttttttaatgaggaatttcatgatgattta 735
            |       |||||||  |  | || | || || | ||||
Db   4913 TTTTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 4873


RESULT 12
X33184/c
ID   X33184 standard; DNA; 7996 BP.
XX
AC   X33184;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-Bcl 2-i-hCD 25.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW   crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP04010.
XX
PR   08-SEP-1997;   97JP-0259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell
XX
PS   Example 3; Page 46-49; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
```

```
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the
CC   plasmid pRx-Bcl 2-i-hCD 25, which contains the human Bcl-2 gene, and
CC   is used in an example from the present invention.
XX
SQ   Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other;


  Query Match             3.0%;  Score 69;  DB 20;  Length 7996;
  Best Local Similarity   46.9%;  Pred. No. 8.6e-05;
  Matches  216;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy     275 ttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttctcagac 334
           || ||| |||||    | | |   ||||| ||||  ||   ||||||||||| |
Db    5532 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5473

Qy     335 agacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatttatg 394
             | | |      |    ||       |   | |   | || ||    | ||| |
Db    5472 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5413

Qy     395 gtcagtctactttttagtagaatatttttggatagcgttgacacgatagatcttattcca 454
            |   | |  |||||  |    | |||||   |    ||        |   | || ||
Db    5412 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5353

Qy     455 tacttctttattattgataattttattttcatttttgctttcattattatacatatttt 514
           |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |   | ||||
Db    5352 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5293

Qy     515 ggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaacact 574
             |          ||   ||||||                 ||| || |     |      |
Db    5292 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5233

Qy     575 ccttttttgacatattaaagcctttattccatctctcaagatatattataaaatttattt 634
             ||||||    | ||      ||| ||   | | |     | | || |    ||| |||
Db    5232 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5173

Qy     635 ttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattat 694
           ||||  |||    ||| |   || |||| | ||    || |  ||||        | || |
Db    5172 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5113

Qy     695 ggtacggaacttttttaatgaggaatttcatgatgattta 735
              |       |||||||  |  | || | || || | ||||
Db    5112 TTTTTTTTTTTTTTTTTTTTTTGCAACTGCAAGAGGGTTTA 5072


RESULT  13
T41852/c
ID   T41852 standard; DNA; 9789 BP.
XX
AC   T41852;
XX
DT   20-FEB-1997  (first entry)
XX
DE   cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
XX
KW   Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW   detection; identification; treatment; prevention; parasite; ss.
XX
OS   Plasmodium falciparum MC type.
XX
FH   Key             Location/Qualifiers
FT   CDS             326..9497
FT                   /*tag= a
FT                   /product= Erythrocyte membrane protein
FT   misc_feature    518..520
FT                   /*tag= b
FT                   /transl_except= GTA encodes Tyrosine
FT   misc_feature    656..658
FT                   /*tag= c
FT                   /transl_except= ATT encodes Leucine
FT   misc_feature    2909..2911
FT                   /*tag= d
FT                   /transl_except= AAC encodes Aspartic acid
FT   misc_feature    3461..3463
FT                   /*tag= e
FT                   /transl_except= GAA encodes Glutamine
FT   misc_feature    5546..5548
FT                   /*tag= f
FT                   /transl_except= CCT encodes Arginine
FT   misc_feature    6254..6256
FT                   /*tag= g
FT                   /transl_except= AAT encodes Lysine
FT   misc_feature    6257..6259
FT                   /*tag= h
FT                   /transl_except= ATA encodes Tyrosine
FT   misc_feature    6263..6265
FT                   /*tag= i
FT                   /transl_except= AAC encodes Lysine
FT   misc_feature    6269..6271
FT                   /*tag= j
FT                   /transl_except= TTC encodes Isoleucine
FT   misc_feature    6272..6274
FT                   /*tag= k
FT                   /transl_except= ATA encodes Histidine
FT   misc_feature    6275..6277
FT                   /*tag= l
FT                   /transl_except= ATT encodes Asparagine
FT   misc_feature    6278..6280
FT                   /*tag= m
FT                   /transl_except= GGA encodes Tryptophan
FT   intron          7754..8478
FT                   /*tag= n
XX
PN   WO9633736-A1.
XX
PD   31-OCT-1996.
XX
PF   26-APR-1996;   96WO-US05798.
XX
PR   27-APR-1995;   95US-0430908.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
XX
PI   Baruch DI,  Howard RJ,  Pasloske BL;
XX
DR   WPI; 1996-497376/49.
DR   P-PSDB; W00384.
XX
PT   New Plasmodium falciparum erythrocyte membrane proteins - used to
PT   develop products for the diagnosis, treatment or prevention of
PT   malaria parasite infections
XX
PS   Disclosure; Figure 12; 149pp; English.
XX
CC   A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC   membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC   protein can be used in the treatment or prevention of symptoms of a
CC   malaria parasite infection.  The polypeptides can inhibit, block or
CC   reverse the sequestration of erythrocytes in patients suffering from
CC   malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC   probes and primers to identify a Plasmodium falciparum parasite, the
CC   primers used to generate characteristic amplification patterns from
```

```
CC   different P. falciparum strains. Antibodies specifically
CC   immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC   used in diagnosis of malaria infection. This sequence encodes the
CC   PfEMP1 protein of the MC type of Plasmodium falciparum.  An
CC   alternative, truncated version of the coding sequence (a cDNA clone)
CC   is given in T41853.
XX
SQ   Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T; 0 other;

  Query Match             2.8%;  Score 65.6;  DB 17;  Length 9789;
  Best Local Similarity   46.5%;  Pred. No. 0.00042;
  Matches  212;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy     385 aatatttatggtcagtctacttttttagtagaatatttttggatagcgttgacacgataga 444
           || || | |  |   | |  |||||  ||   | ||| |   |    || |    ||
Db    8308 AAAATATTTTTTTTTTTTTTTTTTTTTTTAATTTTTTTATTTTTTTTTTTTATTTTATTTT 8249

Qy     445 tcttattccatacttctttattattgataatttttattttcatttttttgctttcattatta 504
           | ||| |   |   | ||| | |||  |  |||| ||    ||||||  |||  || |
Db    8248 TTTTAATTTTTTTATTTTTTTAATTTTTTTTTTTTTTTACATTTTTTTTATTTTTTTTTAT 8189

Qy     505 tacatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataac 564
           |  |||| |   |  | |    ||      ||||   | |        |||||| ||
Db    8188 TTTATATATATATATATATTTTTTTTTAACATTTTTTTAAATTTTTTTTTTTTATTTTATA 8129

Qy     565 actaaacactccttttttgacatattaaagcctttattccatctctcaagatatattata 624
           |   |   |   ||||     | ||      |||||   || | |  | |||| || ||
Db    8128 ATATATATATTTCTTTTAACATTTTTTTTTATTTTATGATATATATATATATATTTTTTA 8069

Qy     625 aaatttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaac 684
           | |||| ||||| |  |||    | | | |   ||||| | |||  || |   |   ||
Db    8068 ACATTTTTTTTTCTTTTTTTTTTATTTAATAAATTTTTTTTTATTTTATTTCTAATAAA 8009

Qy     685 gagctattatggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttc 744
           | ||||  |    |  ||| || |   |    |||  | || |  |      ||||
Db    8008 TTTTTTTTATAATCATTATTTTTATTAATATAATTTTTTTTTTATAAAATTTTTCATTTC 7949

Qy     745 tctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccat 804
           | | ||   | |  ||   | |  | |      |    |  |    || | |  | |
Db    7948 TTTTTTTTTTATGAATTTTTTTTTTTTTATTTATTTTAATAAATTTTTTTTATATTTCTT 7889

Qy     805 ttaaaaactgaaaatattttaaaattatttgtctat 840
           ||      |    || ||||  | || ||| | |||
Db    7888 TTTTCTTTTTTTCATTTTTTTCATTTTTTTCTTTAT 7853


RESULT  14
A70246/c
ID   A70246 standard; DNA; 2364 BP.
XX
AC   A70246;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:379.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 558-559; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 2364 BP; 1222 A; 247 C; 213 G; 682 T; 0 other;

  Query Match             2.8%;  Score 64.6;  DB 21;  Length 2364;
  Best Local Similarity   47.5%;  Pred. No. 0.0005;
  Matches  301;  Conservative    0;  Mismatches  315;  Indels   18;  Gaps    3;

Qy     353 tagcattggaaaatacatatcactattcttggaatatttatggtcagtctacttttttagt 412
           ||   |||||||||||||   | ||||| |   ||    ||  |   |     |  || |
Db    2037 TACATTTGGAAAATACATTGGATTATTCATATTATTGACATTATTCATACTAATAATATT 1978

Qy     413 agaatatttttggatagcgttgacacgatagatcttattccatacttctttattattgat 472
           |       | | ||||   || | || || ||| |||||       |    | |||||||
Db    1977 ATTCATACTATTGATATTATTCATACTATTGATATTATTGACATTATTCACACTATTGAT 1918

Qy     473 aattttattttcatttttttgctttcattattatacatattttg------gtggagaaagag 526
           | | ||  |||  ||||||   ||  |  || |   | ||||       ||    |
Db    1917 ATTATTTATTTTTTTTTTTTTTATTGGTGTTTTTCTTTTTTTTTCCCTTTGTAACTATTTC 1858

Qy     527 gttgggctttttttgaaagagacaaaaatttattataacactaaacactccttttttgaca 586
           || |  || | || | ||||  |    |   |||| |   | |||| |    ||||   |
Db    1857 ATTTGAATTATCTGCATGAGATGACTTTAATTTATTATCATCAACAATAACATTTTTTGA 1798

Qy     587 tattaaagcctttattccatctctcaagatatattataa---------aatttatttttt 637
           |||    |  ||  ||  ||    ||  |||  | |           |||||  ||||||
Db    1797 TATATGTGAATTATTTTTATTATCCATATTATTATTTGTGTTTATATTAATTTGTTTTTT 1738

Qy     638 taatttaagatttctgaattattttatcttaaattgtgattttaaacgagctattatggt 697
           |||  |  |||  |    |  ||| | ||   |  |  |||  ||          ||  |
Db    1737 TAAGTGACTATTATTTTTATTATTTTTATTGTTTGATTGTTTCGAAACTTGAGACATATT 1678

Qy     698 acggaacttttttttaatgaggaatttcatgatgatttaggaattttctctcttggaaaaag 757
             ||    | |||  |  || | | | ||||| |||| |    ||  | ||||   | |
Db    1677 TTGGTTGGTATTTATAATAGAATTATTATGATCATTTTG---ATTAATATCTTTATATAA 1621

Qy     758 gcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccatttaaaaactgaaa 817
           | |    | | || |  || ||          | ||| ||| |    |||      | |
Db    1620 ACATTTGTTATTATTATTATTATTATTATTATTAAAATGTATATTTTTTTTTTCTTTATT 1561
```

```
Qy    818 atattttaaaattatttgtctatattctaaattgagctttggatcaaactttaggccagg 877
          | ||||   ||| | | | ||  |||   |   |  |||| ||||   ||||   |
Db   1560 TTTTTTTTTCATTTTATTTGTAGGTTCATTAAAAAAATTTGTATCATTATTTACTTTAAA 1501

Qy    878 accagctcatgcgttctcattcttcctttctcactctttctctcancactcacctctgt 937
          |  |         |  | |||  |   |   ||  | | | | | |  | | |  |   |
Db   1500 AGTATTATTATGTTGTTGATTTATTGGTGATTCCATGTGTATATTAAAATTTATATTGTT 1441

Qy    938 attcattctgttgtttgggatagaaaaatcataa 971
          ||||||  |    | ||    ||| | ||  | |
Db   1440 GTTCATTTTTAATTGTGTATAAGATATATTGTTA 1407


RESULT 15
C28768
ID   C28768 standard; cDNA; 76 BP.
XX
AC   C28768;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 32843.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 32843; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 76 BP; 25 A; 12 C; 17 G; 22 T; 0 other;


  Query Match             2.7%;  Score 63;  DB 21;  Length 76;
  Best Local Similarity   100.0%;  Pred. No. 0.00055;
  Matches   63;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1615 ccttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccat 1674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ccttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccat 60

Qy   1675 ggc 1677
          |||
Db     61 ggc 63


Search completed: April 30, 2001, 19:06:41
Job time: 36815 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:31:27 ; Search time 330.07 Seconds
(without alignments)
1230.952 Million cell updates/sec

Title: US-09-721-183-3
Perfect score: 2327
Sequence: 1 gatggatgcatctcaaaatg.....actcaattctggttttcccc 2327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	305	13.1	333	3	US-09-284-782-36	Sequence 36, Appl
c 2	70.8	3.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
3	59	2.5	5852	1	US-07-867-106-2	Sequence 2, Appli
4	57.4	2.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
c 5	52	2.2	6243	2	US-09-056-075-1	Sequence 1, Appli
c 6	49.6	2.1	2223	1	US-08-257-073-4	Sequence 4, Appli
7	48.4	2.1	731	1	US-08-451-405A-2	Sequence 2, Appli
c 8	46.8	2.0	1511	1	US-07-991-867B-8	Sequence 8, Appli
c 9	46.8	2.0	1511	1	US-08-107-755A-8	Sequence 8, Appli
c 10	46.8	2.0	1511	2	US-08-544-332-8	Sequence 8, Appli
c 11	46.4	2.0	2230	4	US-08-378-313-24	Sequence 24, Appl
c 12	46	2.0	10395	1	US-08-245-809-3	Sequence 3, Appli
c 13	46	2.0	10396	1	US-08-245-809-5	Sequence 5, Appli
c 14	46	2.0	10798	1	US-08-107-748-2	Sequence 2, Appli
c 15	46	2.0	10798	5	PCT-US92-01385-2	Sequence 2, Appli
c 16	46	2.0	10965	1	US-08-107-748-4	Sequence 4, Appli
c 17	46	2.0	10965	5	PCT-US92-01385-4	Sequence 4, Appli
c 18	45.8	2.0	4838	4	US-08-852-629-15	Sequence 15, Appl
c 19	45.8	2.0	6152	4	US-08-973-462-1	Sequence 1, Appli
20	45.4	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
21	44.8	1.9	2960	3	US-08-913-842-3	Sequence 3, Appli
22	44.2	1.9	1511	1	US-07-991-867B-8	Sequence 8, Appli
23	44.2	1.9	1511	1	US-08-107-755A-8	Sequence 8, Appli
24	44.2	1.9	1511	2	US-08-544-332-8	Sequence 8, Appli
c 25	44.2	1.9	2447	2	US-09-014-969-14	Sequence 14, Appl
26	44	1.9	12124	1	US-08-181-271A-36	Sequence 36, Appl
27	44	1.9	12124	1	US-08-449-315-36	Sequence 36, Appl
28	44	1.9	12124	1	US-08-444-803-36	Sequence 36, Appl
29	44	1.9	12124	1	US-08-449-043-36	Sequence 36, Appl
30	44	1.9	12124	1	US-08-456-265A-36	Sequence 36, Appl
31	44	1.9	12124	1	US-08-455-416-36	Sequence 36, Appl
32	44	1.9	12124	1	US-08-455-244-36	Sequence 36, Appl
33	44	1.9	12124	1	US-08-454-876-36	Sequence 36, Appl
34	44	1.9	12124	2	US-08-457-364-36	Sequence 36, Appl
35	44	1.9	12124	2	US-08-456-262-36	Sequence 36, Appl
36	44	1.9	12124	2	US-08-456-240-36	Sequence 36, Appl
37	44	1.9	12124	2	US-08-455-736-36	Sequence 36, Appl
38	44	1.9	12124	2	US-08-971-217-36	Sequence 36, Appl
39	43.8	1.9	2030	2	US-08-705-937-7	Sequence 7, Appli
c 40	43.8	1.9	2570	2	US-09-056-075-2	Sequence 2, Appli
c 41	43.8	1.9	3659	1	US-08-220-151-72	Sequence 72, Appl
c 42	43.8	1.9	3659	1	US-08-303-124-12	Sequence 12, Appl
c 43	43.8	1.9	3659	1	US-08-413-118-72	Sequence 72, Appl
c 44	43.8	1.9	3659	1	US-08-224-657-48	Sequence 48, Appl
c 45	43.8	1.9	3659	2	US-08-184-009-77	Sequence 77, Appl

ALIGNMENTS

```
RESULT   1
US-09-284-782-36
; Sequence 36, Application US/09284782
; Patent No. 6057111
;  GENERAL INFORMATION:
;    APPLICANT:  ENTERPRISES, LTD., QBI
;    APPLICANT:  Deiss, Louis P.
;    APPLICANT:  Yehiely, Fruma
;    APPLICANT:  Efimova, Elena
;    APPLICANT:  Vasquez-Iaslop, No. 6057111a C.
;    APPLICANT:  Einat, Paz
;    TITLE OF INVENTION:  GENE IDENTIFICATION METHOD
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 6057111thwestern Highway, Suite 410
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/284,782
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Montgomery, Ilene N.
;      REGISTRATION NUMBER:  38,972
;      REFERENCE/DOCKET NUMBER:  0168-00022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (248) 539-5050
;      TELEFAX:  (248) 539-5055
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-284-782-36


  Query Match           13.1%;  Score 305;  DB 3;  Length 333;
  Best Local Similarity  99.7%;  Pred. No. 5.2e-58;
  Matches  316;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy   1887 ggactttggagggcagtgttatttttcccaaagaaagacggccaagggcagaggcatggat 1946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 GGACTTTGGAGGGCAGTGTTATTTTCCCAAAGAAAGACGGCCAAGGGCAGAGGCATGGAT 64

Qy   1947 tctttgcagagcacttccttttggtttttcagtactgtttcatagacagtgggctcacat 2006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 TCTTTGCAGAGCACTTCCTTTTGGTTTTTCAGTACTGTTTCATAGACAGTGGGCTCACAT 124

Qy   2007 gttcctgatagtgctgcagttgcttagaaagcatcccagtt-attgcagtaattagaact 2065
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    125 GTTCCTGATAGTGCTGCAGTTGCTTAGAAAGCATCCCAGTTAATTGCAGTAATTAGAACT 184

Qy   2066 tctggaatatgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattc 2125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 TCTGGAATATGCTAGGGCAGAAGTATGTCAAGTATGTCACATGAAGAAAATGTGAAATTC 244

Qy   2126 aagagtaatccacacgtgagaaactagacaatgtacattcatgtgttctcttgaaaggaa 2185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 AAGAGTAATCCACACGTGAGAAACTAGACAATGTACATTCATGTGTTCTCTTGAAAGGAA 304

Qy   2186 agggagagctgtaagct 2202
          |||||||||||||||||
Db    305 AGGGAGAGCTGTAAGCT 321
```

RESULT 2
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.0%; Score 70.8; DB 2; Length 19124;
Best Local Similarity 48.9%; Pred. No. 1.1e-06;
Matches 256; Conservative 0; Mismatches 257; Indels 11; Gaps 2;

```
Qy    163 ttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgacaagtattacttttt 222
          ||| ||| ||| |      || ||| || | ||    || |      | |  .|||
Db  15920 TTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTT 15861

Qy    223 tgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaataggttctttat 282
          |     |    | | | |       ||   || | | |   |   |||||  || ||| |
Db  15860 TTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTTATTTAAATAAATTTTTTTT 15801

Qy    283 ttttggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatag 342
          |||    | |||   ||||| ||  ||||   ||  |||||||| |  |
Db  15800 ATTTTATGTATATATTTTTTTTTTAACATTTTTTTAATTTTTTTT--------TATTTTA 15749

Qy    343 tagtaacgactagcattggaaaatacatatcactattcttggaatatttatggtcagtct 402
          | ||   | |   |||  || |||  | |  || || ||    | ||||||  |   | |
Db  15748 TGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTTTTGTTTTTATGATATATAT 15689

Qy    403 acttttagtagaatatttttggatagcgttgacacgatagatcttattccatac---tt 459
          |||||  |  ||| |||||   |  | |         |  || || ||    |    ||
Db  15688 TTTTTTTTTTTAATGTTTTTTTTTTTCTTCTTTGTTTTTATTTTTTTTATAATCATTT 15629

Qy    460 ctttattattgataatttttattttcatttttgctttcattattatacatattttggtgg 519
          |||  ||||  | ||||||  |||| |||||||  |   | | ||  | |||  ||||  |
Db  15628 TTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATCTTTTTCATTTTTTATTCT 15569

Qy    520 agaagaggttgggctttttgaaagagacaaaaatttattataacactaaacactccttt 579
          |  | |  ||   ||| ||  ||     |  | ||| | | ||   |     ||  |||
Db  15568 ATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTAAAAAATTTTTCTCCTTTTTT 15509

Qy    580 tttgacatattaaagcctttattccatctctcaagatatattataaaatttatttttttta 639
          |||   | ||      |||| |  || | |    |||| | ||     || ||||  |
Db  15508 TTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTATATTTCATTTTTCTTTTTTTCATT 15449

Qy    640 atttaagatttctgaattattttatcttaaattgtgattttaaa 683
          |   | |||| |   |   ||| | ||  | | | | |  | |
Db  15448 TTAATAAATTTGTTTTTATATTTCTTTTTTAATATAAATACATA 15405
```

RESULT 3
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:

```
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             2.5%;  Score 59;  DB 1;  Length 5852;
  Best Local Similarity   46.3%;  Pred. No. 0.0003;
  Matches  194;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy     266 acaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttttt 325
           | ||| |  |    || || |      | | || | || ||||  ||   |||| |  |
Db    1590 ATAAAAAACTAACCTAGTTATAATTACTTTCCCCTCTTTTTTTTTTTTTTTTTTGTCATG 1649

Qy     326 ttctcagacagacatagtagtaacgactagcattggaaaatacatatcactattcttgga 385
                 | ||  | ||   |    ||  |||| | | |  |  ||  |   |
Db    1650 ACACTTTTTTTTTTTTGTCATGACACTTTTTTTTTAAAAAAAAAAAAAAAAAATGTTAAAA 1709

Qy     386 atatttatggtcagtctactttttagtagaatatttttggatagcgttgacacgatagat 445
             | | || | |   |  |||||  |||  | |||||  |||||   | | |  | |
Db    1710 TACTATTTGATGACATTCATTTTTCCTAGTTTTTTTTTAGATAGATATAAAAATAAATTG 1769

Qy     446 cttattccatacttctttattattgataattttattttcatttttgctttcattattat 505
           | |||       | ||| |||  | |  |||  ||  |  |||  |  ||     ||| |
Db    1770 CCTATCGATATATACTTAATTTATTAAGATTGAATAATATTTTAATTTTTAATAAATTCT 1829

Qy     506 acatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattataaca 565
           || | ||||  |          ||        ||| |||           ||||||| | |
Db    1830 ACTTTTTTTTTTTTTTTCTTTTTTTTTTAAATTTTAAAATTTTTTTTTTTTTATTAGATCT 1889

Qy     566 ctaaacactccttttttgacatattaaagcctttattccatctctcaagatatattataa 625
           |  ||      |   |   | ||||||   | | ||  |      ||| |  ||| | ||
Db    1890 CATAATTAAAAATCAATTTAAAATTAAAAGTTATTTTTAAATATGCAAAAACTATAAAAA 1949

Qy     626 aatttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaac 684
           | |    |  |||||   |   |||   | || |||| | ||   || |  ||||  ||
Db    1950 ACTAATGTAGTTTAACCAACTTTTTTCTATTTCTTTTTTTTTTTTTTTTTTTTTTTAC 2008


RESULT  4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             2.5%;  Score 57.4;  DB 2;  Length 19124;
  Best Local Similarity   43.0%;  Pred. No. 0.00088;
  Matches  280;  Conservative    0;  Mismatches  371;  Indels    0;  Gaps    0;

Qy     265 cacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttttt 324
           ||   || | |  | ||| | |||    ||    | |   |  | ||   ||   | |
Db   17458 CATCTATGGATGATATATATGTGGATCATAATAATGTGACTAATAATAATATGGATGTAC 17517

Qy     325 tttctcagacagacatagtagtaacgactagcattggaaaatacatatcactattcttgg 384
           | || | |   |||| | | | |  | |   | |   ||| |        |      ||
Db   17518 CTACTAAAATGCACATCGAAATGAATATTGTTAATAATAAAAAGGAGATTTTCGAAGAGG 17577

Qy     385 aatatttatggtcagtctactttttagtagaatatttttggatagcgttgacacgataga 444
           |||||      ||||     |      |  || |||  |  |    ||     |   |
Db   17578 AATATCCTATATCAGATATATGGAATATCTAAAATTAATATACTTTTTTTGTGTGTGTCA 17637

Qy     445 tcttattccatacttctttattattgataattttattttcatttttgctttcattatta 504
           | |   |      || ||| |   || |  | || | |||| ||  |||  ||||
Db   17638 TATATATTTTGTATTATTTGTATATGTTTTTATTTTATTTATTTATTTATTTATTTATTG 17697

Qy     505 tacatattttggtggagaagaggttgggctttttgaaagagacaaaaatttattataac 564
           |   | | |  | |  || |   ||   || ||| ||       |   |  || |||
Db   17698 TTTTTGGTATATTTGTAAAAAATATGTTTTTGTTTATAATCATATTATTATATTTTTAAT 17757

Qy     565 actaaacactccttttttgacatattaaagcctttattccatctctcaagatatattata 624
           | |   ||      ||||    | |        || |   || | |      |||||
Db   17758 AATTTGCAACATGATTTTTTTTTTTCTTTCTTATTGTGTAATTTTTTTCATAATATTTAT 17817

Qy     625 aaatttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaac 684
           | || ||| | | |  | |     |     ||| ||| |||||  | |||     | |
Db   17818 ATATATATATGTATTTTATTTTTTAGTATAATAATTGTATCTATATTTGATTAATAATTA 17877

Qy     685 gagctattatggtacggaacttttttttaatgaggaatttcatgatgatttaggaattttc 744
             |||||||||        ||| |||       ||||  |  | || ||   || |
Db   17878 TGTATATTATGGTTATTTTGTTTCTTTTTCTGTACATTTTTTCGTAATATATATATATAT 17937

Qy     745 tctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgccat 804
             |  | ||  ||     | | ||  | ||    | | |  |  |||  | |    |
Db   17938 ATATATATATAATTCTCTTTTTCTAATATATATATCCTTCTATTTTCGATTTTTTCATTT 17997

Qy     805 ttaaaaactgaaaatattttaaaattatttgtctatattctaaattgagctttggatcaa 864
```

```
Db  17998 TTTTCCAGTATTAATTTATTTATTTATTTGTGATATTTTATAATATATTATTTAAATGTG 18057

Qy    865 actttaggccaggaccagctcatgcgttctcattcttccttttctcactct 915
Db  18058 TATTTATATATGTGTTTTATATATGTGTTTTATTTTTGTTACTCTAATTCT 18108


RESULT   5
US-09-056-075-1/c
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.95238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3770..4013
;      OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;      OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1


  Query Match             2.2%;  Score 52;  DB 2;  Length 6243;
  Best Local Similarity   46.1%;  Pred. No. 0.01;
  Matches  214;  Conservative    0;  Mismatches  245;  Indels    5;  Gaps    1;

Qy      55 taaaaaagagaggagattcccccaaacaacaatatttaattttcttagtaaaaagaataa 114
Db    1654 TAAAAGCTAAAGGAAAAGCAACAAAAAAAGAGGAATTAAATGTATTAAGAAAAAAAATAA 1595

Qy     115 cagaatgcatcgtggcaatccttaagcaacattatctatgtggactgcttaaatcagcaa 174
Db    1594 AAGCCCTTAAAGAAAAAGGCTTTAAAAATAAAGAAATTACTCTAATGTTAGAAGTACCAA 1535

Qy     175 aacaccagaagtttggttaacttgggcaatatgacaagtattacttttgggcaaaacta 234
Db    1534 TAAAAACATTAGAACGTCATATTACGTATATGAAAAAAAATGGGCTTTTATAAAGGCTCA 1475

Qy     235 ctcattaagcaatttctct-----agtgtgtcggacacaaataggttctttatttttggc 289
Db    1474 TTTTTTATATTCTTTTCTTCAAAGATTATATAATATAAAAAAATTTTTTTCAAACTTTAA 1415

Qy     290 atgtatgccttttttattttcattcaatttttttttttttctcagacagacatagtagtaac 349
Db    1414 ATAAAAAATATTTTTATATTTTTTTATTTTTTTATTTTTATATTTTTTTTATTTTTTTATT 1355

Qy     350 gactagcattggaaaatacatatcactattcttggaatatttatggtcagtctacttttt 409
Db    1354 TTTATATTTTTTTATTTTTATATTTTTTTATTTTTATATTTTTTTATTTTTATATTTTTT 1295

Qy     410 agtagaatatttttggatagcgttgacacgatagatcttattccatacttctttattatt 469
Db    1294 TATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATT 1235

Qy     470 gataattttattttcatttttgctttcattattatacatattt 513
Db    1234 TTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTTTTATTTTT 1191


RESULT   6
US-08-257-073-4/c
; Sequence 4, Application US/08257073
; Patent No. 5766597
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  de Taisne, Charles
;    APPLICANT:  Tine, John A.
;    TITLE OF INVENTION:  MALARIA RECOMBINANT POXVIRUS VACCINE
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue, 25th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/257,073
;      FILING DATE:  09-JUN-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/075,783
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/852,305
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,183
;      FILING DATE:  20-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2570
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066 CURTMS
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2223 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-257-073-4
```

Query Match 2.1%; Score 49.6; DB 1; Length 2223;
Best Local Similarity 47.4%; Pred. No. 0.026;
Matches 215; Conservative 0; Mismatches 234; Indels 5; Gaps 2;

```
Qy    269 aataggttctttatttttggcatgtatgccttttttattttcattcaatttttttttttttc 328
          |  ||  ||| | ||  || | || |  | | |  | | |   |||    |  ||  | ||
Db    786 AGTATCTTCGTCATCATTTGTATCTTCGTCATCGTTTGTATCTTCTTCATCGTTAGTATC 727

Qy    329 tcagacagacatagtagtaacgactagcattggaaaatacatatcactattcttggaata 388
          |  | ||    | |||    |  |     ||  |  ||    |||| ||   |   |||
Db    726 TTCGTCATCGTTTGTATCTTCTTCATCATTTACATCATCTTCATCATTAACTTCTTGATA 667

Qy    389 tttatggtcagtctacttttttagtagaatatttttggatagcgttgacacgatagatctt 448
          ||| || |     |  |||| | |   || || |  | |  | ||   | ||    | ||
Db    666 TTTTTGATTGTCGTTATTTTCATTTTGATTTTGTGAGGTTTCTTTTTTATGA-CCTTGTG 608

Qy    449 attccatacttctttattattgataatttttattttcatttttttgctttcattattataca 508
          |||  | |    |  || ||  |  | |       |    | | |  | || | || ||
Db    607 ATTTTAAAGAAATAATTTTTTTTTTTGGTATTAACAAAAGAGTAGGTAGCTTTTTGATTCA 548

Qy    509 tattttggtggagaagaggttgggctttttgaaagagacaaaaatttattataacacta 568
          |   ||  |  |  |   |||   |||||  | | || ||| | ||| || |||||
Db    547 TTAATTCTTTAAATAAGTTTTGATTTTTTTGAATATAGTCAACATTTTCTTTTAACATAT 488

Qy    569 aacactccttttttgacatattaaagcctttattccatctctcaagatatattataaaat 628
                ||||||  |||  |  ||||   | ||    | | | |     | |||||| |
Db    487 TC----ATTTTTTTAACACAAGAAAGTAATAATGGATTATATGATTTGTTTTTATAAATT 432

Qy    629 ttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagc 688
             | ||| | ||||   ||   |  | |||| | | ||||    | | || |
Db    431 AAGTGTTTCTTATTTTGTATACTTTCAATATTCTTTTTTAATGCTTTAATTAATTTCTTA 372

Qy    689 tattatggtacggaacttttttaatgaggaatt 722
          |   || ||| | || ||  |  | |    | ||
Db    371 TGAGATTGTAAGAAATTTACTAAATTTTTCATTT 338
```

RESULT 7
US-08-451-405A-2
; Sequence 2, Application US/08451405A
; Patent No. 5736358
; GENERAL INFORMATION:
; APPLICANT: FASEL, NICOLAS JOSEPH
; APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
; TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
; TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE WEBB LAW FIRM
; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
; CITY: PITTSBURGH
; STATE: PENNSYLVANIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 15219-1818
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISK
; COMPUTER: Midwest Micro 486-50
; OPERATING SYSTEM: DOS
; SOFTWARE: WORDPERFECT 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,405A
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,273
; FILING DATE: 15-JAN-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match 2.1%; Score 48.4; DB 1; Length 731;
Best Local Similarity 46.9%; Pred. No. 0.036;
Matches 183; Conservative 0; Mismatches 206; Indels 1; Gaps 1;

```
Qy    287 ggcatgtatgccttttttattttcattcaatttttttttttttctcagacagacatagtagt 346
          || || ||      ||| ||  ||  ||| | | | |||    |    | | |    |
Db    335 GGTATTTAAAGAAATTTTTTAAAATATTATTATATATCTTTAATTGTGCAAAACACACTT 394

Qy    347 aacgactagcattggaaaatacatatcactattcttggaatatttatggtcagtctactt 406
              |   |  |  |||| |    ||  ||  ||| ||| |  |     || |
Db    395 TTAACACACTCTATTATCTTACAAAGGTTTAAAATTTTAATTTTTTTTATTTAATTATTA 454

Qy    407 tttagtagaatatttttggatagcgttgacacgatagatcttattccatacttctttatt 466
          |||  |  ||||  |||   |    ||   |   |  | || ||     | ||
Db    455 TTTTTTTAAATAAATTTTTTTTAATTTTTTAATTTTTTTTTTTTTTTTTACCATCAACCCC 514

Qy    467 attgataatttttattttcatttttgctttcattattatacatattttggtggagaagag 526
            || || |   | |  ||||| ||  |||  ||||| || |||| |   |    || |
Db    515 TTTAATCAAACAAATAACATTTATT-ATTTATTTATTTTATATATATCAATTAGAAATAA 573

Qy    527 gttgggctttttttgaaagagacaaaaatttattataacactaaacactccttttttgaca 586
              ||  |   |  ||  | ||   | |  |   | ||||  |  |||||||  |||
Db    574 AAATATTTTCCTAATAGTAGTAATAATAATTTCTTTTTAATAAAAATACCTTTTTCTACA 633

Qy    587 tattaaagcctttattccatctctcaagatatattataaaatttatttttttaatttaag 646
          |  | |   ||||||  | | ||   |   |||  ||| ||||||||||  || ||
Db    634 TTATTATTTTTTTTATTATTTTTTTCTTTAATCATTCAAAATTTTATTTTTTTTTTTAAAA 693

Qy    647 atttctgaattattttatcttaaattgtga 676
          |      ||  |||  | |  | | | | |
Db    694 AAAAAAAAACAATTAAAACAAACAATTTAA 723
```

RESULT 8
US-07-991-867B-8/c
; Sequence 8, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794

```
;       REFERENCE/DOCKET NUMBER:  UF114.C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  904-375-8100
;       TELEFAX:  904-372-5800
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1511 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Amsacta moorei entemopoxvirus
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  complement (18..218)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  complement (234..782)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  852..1511
US-07-991-867B-8


  Query Match             2.0%;  Score 46.8;  DB 1;  Length 1511;
  Best Local Similarity   47.1%;  Pred. No. 0.096;
  Matches  274;  Conservative    0;  Mismatches  302;  Indels    6;  Gaps    4;

Qy     264 acacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttt 323
           | || ||||||      ||| ||    |       ||| | |||  ||   |||  |||| 
Db    1294 ATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTT 1235

Qy     324 ttttctcagacagacatagtagt-aacgactagcattggaaaatacatatcactattctt 382
           |||      ||  |  || || | ||    |||   |  | | |||||    |     ||
Db    1234 TTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCACAATTT 1175

Qy     383 ggaatatttatggtcagt-ctacttttttagtagaatatttttggatagcgttgacacgat 441
             ||  | ||| |   ||  || | |    |   ||| ||   ||     | | | |||
Db    1174 AAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTATAAGAT 1115

Qy     442 agatcttattccatacttctttattattgataatttttattttcatttttgctttcatta 501
           | |||| | |   |  | ||| | |||       |  | ||| | ||  |  | |    |
Db    1114 ATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATATTTAAA 1055

Qy     502 ttatacatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattat 561
           ||  | |||| |        || |  ||  |   | ||     | | |  | | |  ||
Db    1054 TTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTATCATAA 995

Qy     562 aacactaaacactccttttttgacatattaaagcctttattccatctctcaagatatatt 621
             | | || |  |   | || |   | |||||    ||| ||  || | | || ||  ||
Db     994 TATATTATATGCAATTCTTCT---AAATTAACTAATTTTTTTAATATATTAATATTAATA 938

Qy     622 ataaaatttatttttttaatttaagatttctgaattattttatcttaaattgtgatttta 681
           | |  || | | ||| | | |||   ||| | |||||||  ||| ||||| | |   || |
Db     937 ACATTATCTCTGTTTATTA-TTATTTTTTTTAAATTATAATATTTTAAAATATTTATTAA 879

Qy     682 aacgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatt 741
           ||  |  |   |  ||  ||   |||| || |    ||  |  |  ||   | |
Db     878 AATTATATCAGAATTTAGTAAATCCATTTTGATAATTTTATTTTTTTTTCATTGATTAAT 819

Qy     742 ttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgc 801
           || | | |  ||||   ||  |  | ||||     |  ||   |||  |  |
Db     818 TTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATGTCGATAGAATTAATAATTGGT 759

Qy     802 catttaaaaactgaaaatattttaaaattatttgtctatatt 843
           | | |    |||  || |    | ||||  |     | |||
Db     758 CCTATGTTTTCTGGCAAAACAACAGAATTGATGCGAAAAATT 717


RESULT   9
US-08-107-755A-8/c
; Sequence 8, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-107-755A-8


  Query Match             2.0%;  Score 46.8;  DB 1;  Length 1511;
  Best Local Similarity   47.1%;  Pred. No. 0.096;
  Matches  274;  Conservative    0;  Mismatches  302;  Indels    6;  Gaps    4;

Qy     264 acacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttt 323
           | || ||||||      ||| ||    |       ||| | |||  ||   |||  |||| 
Db    1294 ATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTT 1235

Qy     324 ttttctcagacagacatagtagt-aacgactagcattggaaaatacatatcactattctt 382
           |||      ||  |  || || | ||    |||   |  | | |||||    |     ||
Db    1234 TTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCACAATTT 1175

Qy     383 ggaatatttatggtcagt-ctacttttttagtagaatatttttggatagcgttgacacgat 441
             ||  | ||| |   ||  || | |    |   ||| ||   ||     | | | |||
Db    1174 AAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTATAAGAT 1115
```

```
Qy    442 agatcttattccatacttctttattattgataattttatttttcatttttgctttcatta 501
          | |||| | |   |  |  ||| | |||      |  | ||| | ||  |  | |    |
Db   1114 ATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATATTTAAA 1055

Qy    502 ttatacatattttggtggagaagaggttgggctttttgaaagagacaaaaatttattat 561
          ||  | |||| |         || |  ||  |   | ||     | | |  | | |  ||
Db   1054 TTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTATCATAA 995

Qy    562 aacactaaacactccttttttgacatattaaagcctttattccatctctcaagatatatt 621
          | | || |   |  | || |   | |||||    ||| ||  || | | || ||  ||
Db    994 TATATTATATGCAATTCTTCT---AAATTAACTAATTTTTTTAATATATTAATATTAATA 938

Qy    622 ataaaatttatttttttaatttaagatttctgaattattttatcttaaattgtgatttta 681
          | |  || | | ||| | | |||   ||| | ||||||  ||| ||||| | |   || |
Db    937 ACATTATCTCTGTTTATTA-TTATTTTTTTTAAATTATAATATTTTAAAATATTTATTAA 879

Qy    682 aacgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatt 741
          ||  |  |   |   ||   ||    |||| || |     ||  |  |   ||    | |
Db    878 AATTATATCAGAATTTAGTAAATCCATTTTGATAATTTTATTTTTTTTTCATTGATTAAT 819

Qy    742 ttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgc 801
          || | | |   ||||    ||   |  | ||||      |  ||    |||  |  |
Db    818 TTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATGTCGATAGAATTAATAATTGGT 759

Qy    802 catttaaaaactgaaaatattttaaaattatttgtctatatt 843
          | | |     |||  || |   | ||||  |     | |||
Db    758 CCTATGTTTTCTGGCAAAACAACAGAATTGATGCGAAAAATT 717


RESULT 10
US-08-544-332-8/c
; Sequence 8, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-544-332-8


  Query Match             2.0%;  Score 46.8;  DB 2;  Length 1511;
  Best Local Similarity   47.1%;  Pred. No. 0.096;
  Matches  274;  Conservative    0;  Mismatches  302;  Indels    6;  Gaps    4;

Qy    264 acacaaataggttctttatttttggcatgtatgccttttttattttcattcaatttttttt 323
          | || ||||||      ||| ||    |       ||| | |||  ||   |||  ||||
Db   1294 ATACTAATAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTT 1235

Qy    324 ttttctcagacagacatagtagt-aacgactagcattggaaaatacatatcactattctt 382
          |||       ||  |  || || | ||     |||    |  | | |||||     |     ||
Db   1234 TTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATATTACATGATTCACAATTT 1175

Qy    383 ggaatatttatggtcagt-ctacttttttagtagaatatttttggatagcgttgacacgat 441
            ||  | ||| |   ||  || | |     |    ||| ||  ||     | | | |||
Db   1174 AAAAATTCTATAGAATGTGGTAGTATAATATTACTTATATTGCTATTTTTGTTATAAGAT 1115

Qy    442 agatcttattccatacttctttattattgataattttatttttcatttttgctttcatta 501
          | |||| | |   |  |  ||| | |||      |  | ||| | ||  |  | |    |
Db   1114 ATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAATTTAAATTAATAATATTTAAA 1055

Qy    502 ttatacatattttggtggagaagaggttgggctttttttgaaagagacaaaaatttattat 561
          ||  | |||| |       || |  ||  |   | ||     | | |  | | |  ||
Db   1054 TTTGAAATATATAAACTTTTAATATTTTCTGGAATATTATTTAAAATATTATTATCATAA 995

Qy    562 aacactaaacactccttttttgacatattaaagcctttattccatctctcaagatatatt 621
          | | || |   |  | || |   | |||||    ||| ||  || | | || ||  ||
Db    994 TATATTATATGCAATTCTTCT---AAATTAACTAATTTTTTTAATATATTAATATTAATA 938

Qy    622 ataaaatttatttttttaatttaagatttctgaattattttatcttaaattgtgatttta 681
          | |  || | | ||| | | |||   ||| | ||||||  ||| ||||| | |   || |
Db    937 ACATTATCTCTGTTTATTA-TTATTTTTTTTAAATTATAATATTTTAAAATATTTATTAA 879

Qy    682 aacgagctattatggtacggaacttttttaatgaggaatttcatgatgatttaggaatt 741
          ||  |  |   |   ||   ||    |||| || |     ||  |  |   ||    | |
Db    878 AATTATATCAGAATTTAGTAAATCCATTTTGATAATTTTATTTTTTTTTCATTGATTAAT 819

Qy    742 ttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaaattgtgtgc 801
          || | | |   ||||    ||   |  | ||||      |  ||    |||  |  |
Db    818 TTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATGTCGATAGAATTAATAATTGGT 759

Qy    802 catttaaaaactgaaaatattttaaaattatttgtctatatt 843
          | | |     |||  || |   | ||||  |     | |||
Db    758 CCTATGTTTTCTGGCAAAACAACAGAATTGATGCGAAAAATT 717


RESULT 11
US-08-378-313-24/c
; Sequence 24, Application US/08378313
```

```
; Patent No. 6207881
;  GENERAL INFORMATION:
;    APPLICANT:  THEOLOGIS, ATHANASIOS
;    APPLICANT:  SATO, TAKAHIDO
;    TITLE OF INVENTION:  CONTROL OF FRUIT RIPENING THROUGH
;    TITLE OF INVENTION:  GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/378,313
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/862,493
;      FILING DATE:  02-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  29190-20002.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 856-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2230 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  91..1545
US-08-378-313-24


  Query Match             2.0%;  Score 46.4;  DB 4;  Length 2230;
  Best Local Similarity   48.5%;  Pred. No. 0.13;
  Matches  128;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy     448 tattccatacttctttattattgataattttattttcatttttgctttcattattatac 507
           |||    |  || ||| || ||  |  |||| ||||  ||||||  |||  || || |
Db    2217 TATAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2158

Qy     508 atattttggtggagaagaggttgggctttttgaaagagacaaaaatttattataacact 567
           |   || |   |||  | | |   | |  | |           ||| || |   |
Db    2157 TTTAATTAGCTAAGACAAATTATGATGTCTAAGTACATTTTTTTTTTTTTTTTTTTAAC 2098

Qy     568 aaacactccttttttgacatattaaagcctttattccatctctcaagatatattataaaa 627
           | |   ||  |||| |   |||||    | |  | | | | | |    | || |
Db    2097 ATAACTTCACTTTTGCATTCATTAATTAGCTAAGACAAATTATGATGTCTAAGTACATTT 2038

Qy     628 tttatttttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgag 687
           ||| |||||||  |||   |  |        ||||   || | |  | |  | | || |
Db    2037 TTTTTTTTTTTTTTTTTTAACATAACTTCACTTTTGCATTCATTAATTAGCTAAGACAAA 1978

Qy     688 ctattatggtacggaacttttttt 711
           ||| |||     | || ||||||
Db    1977 TTATGATGTCTAAGTACATTTTTT 1954


RESULT  12
US-08-245-809-3/c
```

```
; Sequence 3, Application US/08245809
; Patent No. 5665579
;  GENERAL INFORMATION:
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  U.S.A.
;      ZIP:  60603-4277
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/245,809
;      FILING DATE:  17-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/771,331
;      FILING DATE:  03-OCT-1991
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  51514
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-552-1311
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2166
;      OTHER INFORMATION:  /note= "N represents approximately
;      OTHER INFORMATION:  0.2 kb of missing sequence in the repetitive DNA
;      OTHER INFORMATION:  region."
;    FEATURE:
;      NAME/KEY:  prim_transcript
;      LOCATION:  3117..7042
US-08-245-809-3


  Query Match             2.0%;  Score 46;  DB 1;  Length 10395;
  Best Local Similarity   46.5%;  Pred. No. 0.23;
  Matches  148;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy     405 tttttagtagaatatttttggatagcgttgacacgatagatcttattccatacttcttta 464
           ||||   |    | ||  |  ||  |  |      || |   |||| |  ||||||  ||
Db    2943 TTTTATTTTACTTCTTCATTTATTTCACTTCATTGAAATGACTTAATGTATACTTTATTT 2884

Qy     465 ttattgataatttttattttcatttttgctttcattattatacatattttggtggagaag 524
           | || | || | |  ||| | ||  ||     | |||  | |   |  |||  |  | ||
Db    2883 TAATCGTTATTATATTTTACTTTGATTAAAATGATTTATTTTGGTTCTTTATTTAACAAA 2824

Qy     525 aggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttttttga 584
           |  |   ||||| |     | |  ||||  | |||| |    |   |     | ||| |
```

```
Db   2823 ATTTACTACTTTTCTTGCTCAAAATAAAAACTGTTATTAGGTCAGTAAAATAATATTTAA 2764

Qy    585 catattaaagcctttattccatctctcaagatatattataaaatttatttttttaattta 644
          | |  |||   |  |||   |     ||  || |  | || | |  |   ||| ||  |
Db   2763 AAGAGAAAATAATAAATTTTGTTGGAGAAATTAAAAAAAAATAGTATTAGATTTTATGGA 2704

Qy    645 agatttctgaattatttttatcttaaattgtgattttaaacgagctattatggtacggaac 704
          |||||| |   | |   ||    || |     |||   || |   |  |   |    |
Db   2703 AGATTTTTCTTTGAAGGTAAAAAAACTCTATGTTTAGGACAAAAGACCAAAATGACCACT 2644

Qy    705 ttttttaatgaggaatt 722
          |||| |||   |   |||
Db   2643 TTTTATTATACATTTATT 2626


RESULT 13
US-08-245-809-5/c
; Sequence 5, Application US/08245809
; Patent No. 5665579
;  GENERAL INFORMATION:
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  U.S.A.
;      ZIP:  60603-4277
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/245,809
;      FILING DATE:  17-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/771,331
;      FILING DATE:  03-OCT-1991
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  51514
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-552-1311
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10396 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon pimpinellifolium
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2166
;      OTHER INFORMATION:  /note= "N represents approximately
;      OTHER INFORMATION:  0.2 kb of missing sequence in the repetitive DNA
;      OTHER INFORMATION:  region."
;    FEATURE:
;      NAME/KEY:  prim_transcript
;      LOCATION:  3117..7043

US-08-245-809-5


  Query Match             2.0%;  Score 46;  DB 1;  Length 10396;
  Best Local Similarity   46.5%;  Pred. No. 0.23;
  Matches  148;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    405 ttttagtagaatatttttggatagcgttgacacgatagatcttattccatacttcttta 464
          ||||   |    | ||  |  ||  |  |     || |   |||| |  ||||||  ||
Db   2943 TTTTATTTTACTTCTTCATTTATTTCACTTCATTGAAATGACTTAATGTATACTTTATTT 2884

Qy    465 ttattgataatttttattttcatttttgctttcattattatacatattttggtggagaag 524
          | || | || | |  ||| | ||  ||    | |||  | |   |  |||  |  | ||
Db   2883 TAATCGTTATTATATTTTACTTTGATTAAAATGATTTATTTTGGTTCTTTATTTAACAAA 2824

Qy    525 aggttgggctttttgaaagagacaaaaatttattataacactaaacactccttttttga 584
          |  |    ||||| |     | |  ||||  | |||| |    |   |     | ||| |
Db   2823 ATTTACTACTTTTCTTGCTCAAAATAAAAACTGTTATTAGGTCAGTAAAATAATATTTAA 2764

Qy    585 catattaaagcctttattccatctctcaagatatattataaaatttatttttttaattta 644
          | |  |||   |  |||   |     ||  || |  | || | |  |   ||| ||  |
Db   2763 AAGAGAAAATAATAAATTTTGTTGGAGAAATTAAAAAAAAATAGTATTAGATTTTATGGA 2704

Qy    645 agatttctgaattatttttatcttaaattgtgattttaaacgagctattatggtacggaac 704.
          |||||| |   | |   ||    || |     |||   || |   |  |   |    |
Db   2703 AGATTTTTCTTTGAAGGTAAAAAAACTCTATGTTTAGGACAAAAGACCAAAATGACCACT 2644

Qy    705 ttttttaatgaggaatt 722
          |||| |||   |   |||
Db   2643 TTTTATTATACATTTATT 2626


RESULT 14
US-08-107-748-2/c
; Sequence 2, Application US/08107748
; Patent No. 5576428
;  GENERAL INFORMATION:
;    APPLICANT:  Fitzmaurice, Leona
;    APPLICANT:  Mirkov, Erik
;    APPLICANT:  Butler, William O.
;    APPLICANT:  Dickinson, Craig D.
;    APPLICANT:  Elliott, Kathryn J.
;    APPLICANT:  Konno, Yoshihiro
;    APPLICANT:  Holtz, Gregory C.
;    TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,748
;      FILING DATE:  20-AUG-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,344
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/771,331
;      FILING DATE:  04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  P51 9503
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  619-546-4737
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10798 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Lycopersicon esculentum
;     FEATURE:
;       NAME/KEY:  precursor_RNA
;       LOCATION:  3520..7445
US-08-107-748-2


  Query Match             2.0%;  Score 46;  DB 1;  Length 10798;
  Best Local Similarity   46.5%;  Pred. No. 0.23;
  Matches  148;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    405 tttttagtagaatatttttggatagcgttgacacgatagatcttattccatacttcttta 464
          ||||  |    | ||  |  ||  |  |     || |  |||| |  ||||||  ||
Db   3346 TTTTATTTTACTTCTTCATTTATTTCACTTCATTGAAATGACTTAATGTATACTTTATTT 3287

Qy    465 ttattgataatttttattttcatttttttgctttcattattatacatattttggtggagaag 524
          | || | || | |  ||| | ||  ||    | |||  | |   |  |||  |  | ||
Db   3286 TAATCGTTATTATATTTTACTTTGATTAAAATGATTTATTTTGGTTCTTTATTTAACAAA 3227

Qy    525 aggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttttttga 584
          |  |    ||||| |     | |  |||| | |||| |    |   |     | ||| |
Db   3226 ATTTACTACTTTTCTTGCTCAAAATAAAAACTGTTATTAGGTCAGTAAAATAATATTTAA 3167

Qy    585 catattaaagcctttattccatctctcaagatatattataaaatttatttttttaattta 644
          | |  |||   |  |||   |     ||  || |  | || | |  |   ||| ||  |
Db   3166 AAGAGAAAATAATAAATTTTGTTGGAGAAATTAAAAAAAAATAGTATTAGATTTTATGGA 3107

Qy    645 agatttctgaattatttttatcttaaattgtgattttaaacgagctattatggtacggaac 704
          ||||||  |   | |   ||    || |     |||   || |   |  |   |    |
Db   3106 AGATTTTTCTTTGAAGGTAAAAAAACTCTATGTTTAGGACAAAAGACCAAAATGACCACT 3047

Qy    705 ttttttttaatgaggaatt 722
          ||||  |||   |   |||
Db   3046 TTTTATTATACATTTATT 3029


RESULT  15
PCT-US92-01385-2/c
; Sequence 2, Application PC/TUS9201385
;  GENERAL INFORMATION:
;     APPLICANT:  Fitzmaurice, Leona
;     APPLICANT:  Mirkov, Erik
;     APPLICANT:  Butler, William O.
;     APPLICANT:  Dickinson, Craig D.
;     APPLICANT:  Elliott, Kathryn J.
;     APPLICANT:  Konno, Yoshihiro
;     APPLICANT:  Holtz, Gregory C.
;     TITLE OF INVENTION:  NOVEL INVERTASE GENE(S) AND USES THEREOF
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fitch, Even, Tabin & Flannery
;       STREET:  135 South LaSalle Street, Suite 900
;       CITY:  Chicago
;       STATE:  IL
;       COUNTRY:  U.S.A.
;       ZIP:  60603-4277
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US92/01385
;       FILING DATE:  19920221
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/660,344
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/771,331
;       FILING DATE:  04-OCT-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie L.
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  52498PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-552-1311
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10798 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Lycopersicon esculentum
;     FEATURE:
;       NAME/KEY:  precursor_RNA
;       LOCATION:  3520..7445
PCT-US92-01385-2


  Query Match             2.0%;  Score 46;  DB 5;  Length 10798;
  Best Local Similarity   46.5%;  Pred. No. 0.23;
  Matches  148;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    405 tttttagtagaatatttttggatagcgttgacacgatagatcttattccatacttcttta 464
          ||||  |    | ||  |  ||  |  |     || |  |||| |  ||||||  ||
Db   3346 TTTTATTTTACTTCTTCATTTATTTCACTTCATTGAAATGACTTAATGTATACTTTATTT 3287

Qy    465 ttattgataatttttattttcatttttttgctttcattattatacatattttggtggagaag 524
          | || | || | |  ||| | ||  ||    | |||  | |   |  |||  |  | ||
Db   3286 TAATCGTTATTATATTTTACTTTGATTAAAATGATTTATTTTGGTTCTTTATTTAACAAA 3227

Qy    525 aggttgggctttttttgaaagagacaaaaatttattataacactaaacactccttttttga 584
          |  |    ||||| |     | |  |||| | |||| |    |   |     | ||| |
Db   3226 ATTTACTACTTTTCTTGCTCAAAATAAAAACTGTTATTAGGTCAGTAAAATAATATTTAA 3167

Qy    585 catattaaagcctttattccatctctcaagatatattataaaatttatttttttaattta 644
          | |  |||   |  |||   |     ||  || |  | || | |  |   ||| ||  |
Db   3166 AAGAGAAAATAATAAATTTTGTTGGAGAAATTAAAAAAAAATAGTATTAGATTTTATGGA 3107

Qy    645 agatttctgaattatttttatcttaaattgtgattttaaacgagctattatggtacggaac 704
          ||||||  |   | |   ||    || |     |||   || |   |  |   |    |
Db   3106 AGATTTTTCTTTGAAGGTAAAAAAACTCTATGTTTAGGACAAAAGACCAAAATGACCACT 3047

Qy    705 ttttttttaatgaggaatt 722
          ||||  |||   |   |||
Db   3046 TTTTATTATACATTTATT 3029


Search completed:  May  1, 2001, 00:33:02
Job time:  56526 sec
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:17:28 ; Search time 12176.5 Seconds
(without alignments)
27.933 Million cell updates/sec

Title: US-09-721-183-3
Perfect score: 2327
Sequence: 1 gatggatgcatctcaaaatg.....actcaattctggttttcccc 2327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	841	36.1	842	106	AL535097	AL535097 AL535097
	2	686.4	29.5	691	107	AU117444	AU117444 AU117444
c	3	593.8	25.5	598	113	AW303579	AW303579 xv20g11.x
	4	568	24.4	569	32	AW021592	AW021592 df25f07.y
c	5	546.4	23.5	584	122	AW973691	AW973691 EST385791
c	6	544.8	23.4	602	229	AG025673	AG025673 Homo sapi
c	7	541	23.2	547	103	AI858596	AI858596 wl40c11.x
c	8	527	22.6	542	104	AI953076	AI953076 wp99a05.x
c	9	521	22.4	525	148	BF439110	BF439110 nab59d09.
c	10	520.8	22.4	560	107	AU144274	AU144274 AU144274
	11	499.2	21.5	567	141	BE932151	BE932151 IL3-HT061
c	12	498.8	21.4	571	107	AU144115	AU144115 AU144115
c	13	476	20.5	478	19	AI380343	AI380343 tf94h10.x
c	14	458	19.7	467	103	AI927871	AI927871 wo66c11.x
	15	454.2	19.5	897	145	BF185507	BF185507 601814588
c	16	454	19.5	458	17	AI167168	AI167168 oo08d02.x
c	17	398.2	17.1	478	139	BE769299	BE769299 PM4-FT002
	18	380	16.3	380	7	AA470941	AA470941 nc66a02.r

c	19	376	16.2	384	10	AA653508	AA653508 ag66f01.s
c	20	369	15.9	376	16	AI130713	AI130713 qc16e02.x
	21	364.6	15.7	472	155	R54192	R54192 yg98d11.r1
	22	361.6	15.5	1114	149	BF541304	BF541304 602068985
c	23	352.8	15.2	411	103	AI889468	AI889468 wn04h02.x
c	24	351.2	15.1	411	103	AI889459	AI889459 wn04g02.x
c	25	343	14.7	343	9	AA553731	AA553731 nl13g08.s
	26	332	14.3	346	201	AQ071245	AQ071245 HS_2268_B
	27	330.8	14.2	354	170	BF916544	BF916544 PM0-UT010
	28	322.8	13.9	357	170	BF916547	BF916547 PM0-UT010
	29	304	13.1	322	121	AW889283	AW889283 QV4-NT002
	30	304	13.1	340	161	BE004240	BE004240 CM0-BN010
c	31	288.4	12.4	319	7	AA470973	AA470973 nc66a02.s
c	32	288.2	12.4	378	103	AI925749	AI925749 wo35a04.x
c	33	278	11.9	283	148	BF434227	BF434227 7o71d05.x
	34	275	11.8	851	168	BF701166	BF701166 602128236
	35	267.8	11.5	303	162	BE083867	BE083867 PM4-BT065
	36	231.4	9.9	449	112	AW211387	AW211387 uo80d05.y
	37	227	9.8	331	153	H39863	H39863 yo57b06.r1
	38	226.6	9.7	269	162	BE083767	BE083767 PM4-BT065
c	39	221.2	9.5	259	168	BF742182	BF742182 QV0-BT085
	40	192.4	8.3	912	144	BF161004	BF161004 601767631
c	41	187.8	8.1	250	163	BE148370	BE148370 CM0-HT024
c	42	177.6	7.6	429	147	BF388827	BF388827 UI-R-BS2-
	43	176.6	7.6	713	174	BG245625	BG245625 602358406
c	44	176	7.6	207	23	AI648391	AI648391 tx68g05.x
	45	145.8	6.3	650	166	BE374059	BE374059 601228096

ALIGNMENTS

```
RESULT   1
AL535097
LOCUS       AL535097      842 bp    mRNA            EST       13-FEB-2001
DEFINITION  AL535097 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF008YI01 5
            prime, mRNA sequence.
ACCESSION   AL535097
VERSION     AL535097.1  GI:12798590
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DF008YI01"
                     /clone_lib="LTI_FL013_FBrn1"
                     /dev_stage="pooled tissue from post conception fetuses (20
                     week, 24 week and 26 week)"
                     /lab_host="DH10B"
                     /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector.  Library was constructed by Life
                     Technologies.  Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      249 a    130 c    126 g    337 t
ORIGIN
```

```
  Query Match             36.1%;  Score 841;  DB 106;  Length 842;
  Best Local Similarity   99.9%;  Pred. No. 4e-160;
  Matches  841;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     193 aacttgggcaatatgacaagtattacttttgggcaaaactactcattaagcaatttctc 252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AACTTGGGCAATATGACAAGTATTACTTTTGGGCAAAACTACTCATTAAGCAATTTCTC 60

Qy     253 tagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttttatttttcatt 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TAGTGTGTCGGACACAAATAGGTTCTTTATTTTTGGCATGTATGCCTTTTTTATTTTCATT 120

Qy     313 caatttttttttttctcagacagacatagtagtaacgactagcattggaaaatacatat 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAATTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATAT 180

Qy     373 cactattcttggaatatttatggtcagtctacttttttagtagaatatttttggatagcgt 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CACTATTCTTGGAATATTTATGGTCAGTCTACTTTTTAGTAGAATATTTTTGGATAGCGT 240

Qy     433 tgacacgatagatcttattccatacttctttattattgataattttattttcatttttttg 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGACACGATAGATCTTATTCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTG 300

Qy     493 ctttcattattatacatattttggtggagaagaggttgggctttttttgaaagagacaaaa 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTTTCATTATTATACATATTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAA 360

Qy     553 atttattataacactaaacactccttttttgacatattaaagcctttattccatctctca 612
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTTATTATAACACTAAACACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCA 420

Qy     613 agatatattataaaatttatttttttaatttaagatttctgaattattttatcttaaatt 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGATATATTATAAAATTTATTTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATT 480

Qy     673 gtgattttaaacgagctattatggtacggaacttttttttaatgaggaatttcatgatgat 732
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTGATTTTAAACGAGCTATTATGGTACGGAACTTTTTTTTAATGAGGAATTTCATGATGAT 540

Qy     733 ttaggaattttctctcttggaaaaggcttcccctgtgatgaaaatgatgtgccagctaaa 792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTAGGAATTTTCTCTCTTGGAAAAGGCTTCCCCTGTGATGAAAATGATGTGCCAGCTAAA 600

Qy     793 attgtgtgccatttaaaaactgaaaatattttaaaattatttgtctatattctaaattga 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATTGTGTGCCATTTAAAAACTGAAAATATTTTAAAATTATTTGTCTATATTCTAAATTGA 660

Qy     853 gctttggatcaaactttaggccaggaccagctcatgcgttctcattcttccttttctcac 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCTTTGGATCAAACTTTAGGCCAGGACCAGCTCATGCGTTCTCATTCTTCCTTTTCTCAC 720

Qy     913 tctttctctcancactcacctctgtattcattctgttgtttgggatagaaaaatcataaa 972
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCTTTCTCTCATCACTCACCTCTGTATTCATTCTGTTGTTTGGGATAGAAAAATCATAAA 780

Qy     973 gagccaacccatctcagaacgttgtggattgagagagacactacatgactccaagtatat 1032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GAGCCAACCCATCTCAGAACGTTGTGGATTGAGAGAGACACTACATGACTCCAAGTATAT 840

Qy    1033 ga 1034
           ||
Db     841 GA 842
```

```
RESULT   2
AU117444
LOCUS       AU117444      691 bp    mRNA            EST       19-OCT-2000
DEFINITION  AU117444 HEMBA1 Homo sapiens cDNA clone HEMBA1001411 5', mRNA
            sequence.
ACCESSION   AU117444
```

```
VERSION     AU117444.1  GI:10932413
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1001411"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      211 a     93 c    104 g    280 t      3 others
ORIGIN
```

Query Match 29.5%; Score 686.4; DB 107; Length 691;
Best Local Similarity 99.4%; Pred. No. 7.2e-129;
Matches 687; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      91 taatttttcttagtaaaaagaataacagaatgcatcgtggcaatccttaagcaacattatc 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TAATTTTCTTAGTAAAAAGAATAACAGAATGCATCGTGGCAATCCTTAAGCAACATTATC 60

Qy     151 tatgtggactgcttaaatcagcaaaacaccagaagtttggttaacttgggcaatatgaca 210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TATGTGGACTGCTTAAATCAGCAAAACACCAGAAGTTTGGTTAACTTGGGCAATATGACA 120

Qy     211 agtattacttttgggcaaaactactcattaagcaatttctctagtgtgtcggacacaaa 270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGTATTACTTTTTGGGCAAAACTACTCATTAAGCAATTTCTCTAGTGTGTCGGACACAAA 180

Qy     271 taggttctttatttttggcatgtatgccttttttatttttcattcaatttttttttttttctc 330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TAGGTTCTTTATTTTTGGCATGTATGCCTTTTTTATTTTCATTCAATTTTTTTTTTTTTCTC 240

Qy     331 agacagacatagtagtaacgactagcattggaaaatacatatcactattcttggaatatt 390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGACAGACATAGTAGTAACGACTAGCATTGGAAAATACATATCACTATTCTTGGAATATT 300

Qy     391 tatggtcagtctactttttagtagaatatttttggatagcgttgacacgatagatcttat 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TATGGTCAGTCTACTTTTTAGTAGAATATTTTTGGATAGCGTTGACACGATAGATCTTAT 360

Qy     451 tccatacttctttattattgataatttttattttcatttttgctttcattattatacata 510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCCATACTTCTTTATTATTGATAATTTTATTTTCATTTTTTGCTTTCATTATTATACATA 420

Qy     511 ttttggtggagaagaggttgggctttttgaaagagacaaaaatttattataacactaaa 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTTTGGTGGAGAAGAGGTTGGGCTTTTTTGAAAGAGACAAAAATTTATTATAACACTAAA 480

Qy     571 cactccttttttgacatattaaagcctttattccatctctcaagatatattataaaattt 630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CACTCCTTTTTTGACATATTAAAGCCTTTATTCCATCTCTCAAGATATATTATAAAATTT 540

Qy     631 attttttaatttaagatttctgaattattttatcttaaattgtgattttaaacgagcta 690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     541 ATTTTTTAATTTAAGATTTCTGAATTATTTTATCTTAAATTGTGATTTTAAGCGAGCTA 600

Qy     691 ttatggtacggaacttttttttaatgaggaatttcatgatgatttaggaatttttctctctt 750
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     601 TTATGGTACGGAACTTTTTTTTAATGAGGAATTTCATGATGATTTANGAATTTTCTCTCTT 660

Qy     751 ggaaaaggcttcccctgtgatgaaaatgatg 781
           |||||||||||| ||||| ||||||||||||||
Db     661 GGAAAAGGCTTNCCCTGNGATGAAAATGATG 691
```

```
RESULT   3
AW303579/c
LOCUS       AW303579       598 bp    mRNA            EST       18-JAN-2000
DEFINITION  xv20g11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2813732 3', mRNA sequence.
ACCESSION   AW303579
VERSION     AW303579.1  GI:6713268
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2813732"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      203 a    122 c     92 g    181 t
ORIGIN
```

Query Match 25.5%; Score 593.8; DB 113; Length 598;
Best Local Similarity 99.7%; Pred. No. 3.7e-110;
Matches 595; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    1081 agagtatgcccaattctctctacatgacatattgagatttttttttaatcaacttttaaga 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 AGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTTAATCAACTTTTAAGA 539

Qy    1141 tagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaacaagc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    538 TAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGC 479

Qy   1201 atggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 ATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTC 419

Qy   1261 tttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctctatt 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 TTTCTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATT 359

Qy   1321 atgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatgcctg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 ATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTG 299

Qy   1381 aggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 AGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATT 239

Qy   1441 agcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaact 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACT 179

Qy   1501 gctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccag 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAG 119

Qy   1561 tggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcctttt 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 TGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTT 59

Qy   1621 gccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 1677
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db     58 GCCTTAAGTTGAGAGGAGGTCACCTTTAGCTACTGTCTTTGGTTTGAGAGCCAAGGC 2
```

RESULT 4
AW021592
LOCUS AW021592 569 bp mRNA EST 13-SEP-1999
DEFINITION df25f07.y1 Morton Fetal Cochlea Homo sapiens cDNA clone IMAGE:2484612 5', mRNA sequence.
ACCESSION AW021592
VERSION AW021592.1 GI:5875122
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R. and Morton,C.C.
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening
JOURNAL Genomics 23, 42-50 (1994)
MEDLINE 95130111
COMMENT Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: ccmorton@bics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes of Health Intramural Sequencing Center (NISC; see http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Plate: LLAM6182 row: K column: 13
Seq primer: M13RP1 reverse primer (ABI).
FEATURES Location/Qualifiers
source 1. .569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2484612"
/clone_lib="Morton Fetal Cochlea"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned unidirectionally. Primer: Oligo dT. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 182 a 95 c 94 g 198 t
ORIGIN

```
Query Match 24.4%; Score 568; DB 32; Length 569;
Best Local Similarity 99.8%; Pred. No. 6.1e-105;
Matches 568; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy    828 attatttgtctatattctaaattgagctttggatcaaactttaggccaggaccagctcat 887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATTATTTGTCTATATTCTAAATTGAGCTTTGGATCAAACTTTAGGCCAGGACCAGCTCAT 60

Qy    888 gcgttctcattcttccttttctcactctttctctcancactcacctctgtattcattctg 947
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     61 GCGTTCTCATTCTTCCTTTTCTCACTCTTTCTCTCATCACTCACCTCTGTATTCATTCTG 120

Qy    948 ttgtttgggatagaaaaatcataaagagccaacccatctcagaacgttgtggattgagag 1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGTTTGGGATAGAAAAATCATAAAGAGCCAACCCATCTCAGAACGTTGTGGATTGAGAG 180

Qy   1008 agacactacatgactccaagtatatgagaaaaggacagagctctaattgataactctgta 1067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGACACTACATGACTCCAAGTATATGAGAAAAGGACAGAGCTCTAATTGATAACTCTGTA 240

Qy   1068 gttcaaaaggaaaagagtatgcccaattctctctacatgacatattgagatttttttaa 1127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTCAAAAGGAAAAGAGTATGCCCAATTCTCTCTACATGACATATTGAGATTTTTTTTAA 300

Qy   1128 tcaacttttaagatagtgatgttctgttctaaactgttctgttttagtgaaggtagattt 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCAACTTTTAAGATAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTT 360

Qy   1188 ttataaaacaagcatggggattcttttctaaggtaatattaatgagaagggaaaaaagta 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTATAAAACAAGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTA 420

Qy   1248 tctttaacagctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtg 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCTTTAACAGCTCTTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTG 480

Qy   1308 cacataatctattatgatccaatgcaaatacagctccaaaaatattaaatgtatatatat 1367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CACATAATCTATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATAT 540

Qy   1368 tttaaaatgcctgaggaaatacatttttc 1396
          |||||||||||||||||||||||||||||
Db    541 TTTAAAATGCCTGAGGAAATACATTTTTC 569
```

RESULT 5
AW973691/c
LOCUS AW973691 584 bp mRNA EST 02-JUN-2000
DEFINITION EST385791 MAGE resequences, MAGM Homo sapiens cDNA, mRNA sequence.
ACCESSION AW973691
VERSION AW973691.1 GI:8164874
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 584)

```
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 324
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGM"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      203 a    118 c     85 g    178 t
ORIGIN
```

```
  Query Match             23.5%;  Score 546.4;  DB 122;  Length 584;
  Best Local Similarity   97.1%;  Pred. No. 1.4e-100;
  Matches  567;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy   1093 attctctctacatgacatattgagatttttttttaatcaacttttaagatagtgatgttct 1152
          | ||| | ||||        | || ||||||||||||||| ||||||||||  ||||| |
Db    583 AATCTTTTTACAATGACATATTAGTTTTTTTTTTAATCAAC-TTTAAGATAGGAATGTTTT 525

Qy   1153 gttctaaactgttctgttttagtgaaggtagatttttataaaacaagcatggggattctt 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     524 GTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATGGGGATTCTT 465

Qy   1213 ttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcc 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     464 TTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTTGTTGAAGCC 405

Qy   1273 tgtggtagcacattatgtttataattgcacatgtgcacataatctattatgatccaatgc 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     404 TGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATGATCCAATGC 345

Qy   1333 aaatacagctccaaaaatattaaatgtatatatattttaaaatgcctgaggaaatacatt 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 AAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACATT 285

Qy   1393 tttcttaataaactgaagagtctcagtatggctattaaaataattattagcctcctgttg 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 TTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTTG 225

Qy   1453 tgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgttt 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 TGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTTT 165

Qy   1513 agtaaataaaattaatgcatttctagaggggggaatatctgccatccagtggtggaaatgt 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATGT 105

Qy   1573 ggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgccttaagttga 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104 GGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGA 45

Qy   1633 gaggaggtcaactttagctactgtctttggtttgagagccatgg 1676
          |||||||||||||||||||||||||||||||||||||||||||
Db     44 GAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGG 1
```

```
RESULT   6
AG025673/c
LOCUS       AG025673      602 bp    DNA             GSS       21-DEC-1999
DEFINITION  Homo sapiens DNA, chromosome 8q23, reverse end of BAC clone:
            KB1432F7, genomic survey sequence.
ACCESSION   AG025673
VERSION     AG025673.1  GI:6624364
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Homo sapiens DNA, clone_lib:KEIO BAC library clone:KB1432F7.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     The BAC end sequence of Homo sapiens BAC clone KB1432F7
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 602)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1999) to the DDBJ/EMBL/GenBank databases.
            Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
            Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
            (E-mail:nshimizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370,
            Fax:81-3-3351-2370)
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="KB1432F7"
                     /clone_lib="KEIO BAC library"
                     /map="8q23"
                     /note="This sequence is reverse end of BAC clone
                     KB1432F7."
BASE COUNT      190 a    127 c     87 g    195 t      3 others
ORIGIN
```

```
  Query Match             23.4%;  Score 544.8;  DB 229;  Length 602;
  Best Local Similarity   94.4%;  Pred. No. 3e-100;
  Matches  568;  Conservative    0;  Mismatches   32;  Indels    2;  Gaps    2;

Qy   1138 agatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaca 1197
          ||||||| |||||||||||||||||||||||| |||||||||||||||||||  ||
Db    602 AGATAGTNATGTTCTGTTCTAAACTGTTCTGTTTAAGTGAAGGTAGATTTTTATAAANCA 543

Qy   1198 agcatggggattcttttctaaggtaatattaatgagaaggg-aaaaaagtatctttaaca 1256
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||
Db    542 AGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAAGTATCTTTANCA 483

Qy   1257 gctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatc 1316
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GCTCTTTGTTAAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATC 423

Qy   1317 tattatgatccaatgcaaatacagctccaaaaatattaaatgtatatatatttttaaaatg 1376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 TATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATG 363

Qy   1377 cctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataat 1436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CCTGAGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAAT 303

Qy   1437 tattagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtg 1496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TATTAGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTG 243

Qy   1497 aactgctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccat 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AACTGCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCAT 183

Qy   1557 ccagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcc 1616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCC 123

Qy   1617 ttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatgg 1676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGG 63
```

```
Qy   1677 c-nnnnnnnnnnnnnnnnnnnnnnnnnnnngtcgtctttggtgagccagtaaggtgaaagc 1735
          |                              ||||||||||||||||||||||||||||||
Db     62 CAAAAAAAAAAAAAAGAAAAAAAGATCAAGTCGTCTTTGGTGAGCCAGTAAGGTGAAAGC 3

Qy   1736 tt 1737
          ||
Db      2 TT 1


RESULT   7
AI858596/c
LOCUS       AI858596      547 bp    mRNA            EST       07-MAR-2000
DEFINITION  wl40c11.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2427380 3',
            mRNA sequence.
ACCESSION   AI858596
VERSION     AI858596.1  GI:5512212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1835   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2427380"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      180 a    116 c     81 g    170 t
ORIGIN


  Query Match             23.2%;  Score 541;  DB 103;  Length 547;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-99;
  Matches  541;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1137 aagatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaac 1196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 AAGATAGTGATGTTCTGTTCTAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAAC 488

Qy   1197 aagcatggggattctttttctaaggtaatattaatgagaagggaaaaaagtatctttaaca 1256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 AAGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACA 428

Qy   1257 gctctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatc 1316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 GCTCTTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATC 368

Qy   1317 tattatgatccaatgcaaatacagctccaaaaatattaaatgtatatatattttaaaatg 1376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 TATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATG 308

Qy   1377 cctgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataat 1436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CCTGAGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAAT 248

Qy   1437 tattagcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtg 1496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 TATTAGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTG 188

Qy   1497 aactgctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccat 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 AACTGCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCAT 128

Qy   1557 ccagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcc 1616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 CCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCC 68

Qy   1617 ttttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatgg 1676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 TTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGG 8

Qy   1677 c 1677
          |
Db      7 C 7


RESULT   8
AI953076/c
LOCUS       AI953076      542 bp    mRNA            EST       08-MAR-2000
DEFINITION  wp99a05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2469872 3',
            mRNA sequence.
ACCESSION   AI953076
VERSION     AI953076.1  GI:5745386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1262   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2469872"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
```

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 179 a 116 c 82 g 164 t 1 others
ORIGIN

Query Match 22.6%; Score 527; DB 104; Length 542;
Best Local Similarity 99.6%; Pred. No. 1.2e-96;
Matches 538; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy   1139 gatagtgatgttctgttctaaactgttctgttttagtgaaggtagatttttataaaaacaa 1198
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    542 GATAGTGATGTTCTGTTCTAAACTGTTCTGTTNTAGTGAAGGTAGATTTTTATAAAACAA 483

Qy   1199 gcatggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagc 1258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGC 423

Qy   1259 tctttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatcta 1318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 TCTTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTA 363

Qy   1319 ttatgatccaatgcaaatacagctccaaaaaatattaaatgtatatatattttaaaatgcc 1378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCC 303

Qy   1379 tgaggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataatta 1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TGAGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTA 243

Qy   1439 ttag-cctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtga 1497
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 TTAGCCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGA 183

Qy   1498 actgctgccctgtttagtaaataaaattaatgcatttctagaggggaatatctgccatc 1557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 ACTGCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATC 123

Qy   1558 cagtggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagcct 1617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 CAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCT 63

Qy   1618 tttgccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 1677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGC 3
```

RESULT 9
BF439110/c
LOCUS BF439110 525 bp mRNA EST 19-JAN-2001
DEFINITION nab59d09.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3270185 3', mRNA sequence.
ACCESSION BF439110
VERSION BF439110.1 GI:11451627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 465.
FEATURES Location/Qualifiers
source 1. .525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3270185"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NbHSF pool 1: 309384-310919, 323208-325895 Soares Nb2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares Nb2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NbHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NbHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 172 a 111 c 79 g 163 t
ORIGIN

Query Match 22.4%; Score 521; DB 148; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.9e-95;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1157 taaactgttctgttttagtgaaggtagatttttataaaacaagcatggggattcttttct 1216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    525 TAAACTGTTCTGTTTTAGTGAAGGTAGATTTTTATAAAACAAGCATGGGGATTCTTTTCT 466

Qy   1217 aaggtaatattaatgagaagggaaaaaagtatctttaacagctctttgttgaagcctgtg 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 AAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCTTTGTTGAAGCCTGTG 406

Qy   1277 gtagcacattatgtttataattgcacatgtgcacataatctattatgatccaatgcaaat 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 GTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATGATCCAATGCAAAT 346

Qy   1337 acagctccaaaaatattaaatgtatatatatttttaaaatgcctgaggaaatacatttttc 1396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 ACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACATTTTTC 286

Qy   1397 ttaataaactgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtg 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 TTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTTGTGTG 226

Qy   1457 gctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagta 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 GCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTTTAGTA 166

Qy   1517 aataaaattaatgcatttctagaggggaatatctgccatccagtggtggaaatgtggag 1576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 AATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATGTGGAG 106

Qy   1577 taaagaagctggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagagg 1636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 TAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGG 46

Qy   1637 aggtcaactttagctactgtctttggtttgagagccatggc 1677
          |||||||||||||||||||||||||||||||||||||||||
Db     45 AGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGC 5
```

RESULT 10
AU144274/c

```
LOCUS       AU144274      560 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU144274 HEMBA1 Homo sapiens cDNA clone HEMBA1001411 3', mRNA
            sequence.
ACCESSION   AU144274
VERSION     AU144274.1  GI:11005795
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
            ,S. and Isogai,T.).
  TITLE     HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
            Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
            ,Y., Nagai,T., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1001411"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      186 a    116 c     87 g    168 t      3 others
ORIGIN
```

```
  Query Match             22.4%;  Score 520.8;  DB 107;  Length 560;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-95;
  Matches  547;  Conservative    0;  Mismatches    9;  Indels    2;  Gaps    2;

Qy    1120 ttttttaatcaacttttaagatagtgatgttctgttctaaactgttctgttttagtgaag 1179
           |||||| | | || |||||||||| ||||||| ||||||| ||| ||||||| ||| |||
Db     556 TTTTTTTAACCACCTTTAAGATAG-GATGTTCNGTTCTAACCTG-TCTGTTTAAGTNAAG 499

Qy    1180 gtagatttttataaaacaagcatggggattcttttctaaggtaatattaatgagaaggga 1239
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     498 GTAGATTTTTATAAACCAAGCATGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGA 439

Qy    1240 aaaaagtatctttaacagctctttgttgaagcctgtggtagcacattatgtttataattg 1299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     438 AAAAAGTATCTTTAACAGCTCTTTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTG 379

Qy    1300 cacatgtgcacataatctattatgatccaatgcaaatacagctccaaaaatattaaatgt 1359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     378 CACATGTGCACATAATCTATTATGATCCAATGCAAATACAGCTCCAAAAATATTAAATGT 319

Qy    1360 atatatattttaaaatgcctgaggaaatacatttttcttaataaactgaagagtctcagt 1419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318 ATATATATTTTAAAATGCCTGAGGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGT 259

Qy    1420 atggctattaaaatattattagcctcctgttgtgtggctgcaaaacatcacaaagtgac 1479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 ATGGCTATTAAAATAATTATTAGCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGAC 199

Qy    1480 cggtcttgagacctgtgaactgctgccctgtttagtaaataaaattaatgcatttctaga 1539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 CGGTCTTGAGACCTGTGAACTGCTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGA 139
```

```
Qy    1540 gggggaatatctgccatccagtggtggaaatgtggagtaaagaagctggtggtctgcttc 1599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 GGGGGAATATCTGCCATCCAGTGGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTC 79

Qy    1600 tgtgctgtatgccagccttttgccttaagttgagaggaggtcaactttagctactgtctt 1659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 TGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTT 19

Qy    1660 tggtttgagagccatggc 1677
           ||||||||||||||||||
Db      18 TGGTTTGAGAGCCATGGC 1
```

```
RESULT  11
BE932151
LOCUS       BE932151      567 bp    mRNA            EST       02-OCT-2000
DEFINITION  IL3-HT0618-180800-111-F02 HT0618 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE932151
VERSION     BE932151.1  GI:10458227
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=IL3-HT0618-180
            800-111-F02&t3=2000-08-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 339.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0618"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      181 a     99 c    140 g    146 t      1 others
ORIGIN
```

```
  Query Match             21.5%;  Score 499.2;  DB 141;  Length 567;
  Best Local Similarity   94.1%;  Pred. No. 5e-91;
  Matches  523;  Conservative    0;  Mismatches   31;  Indels    2;  Gaps    2;

Qy    1597 ttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaactttagctactgt 1656
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      12 TCCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGT 71

Qy    1657 ctttggtttgagagccatggc-nnnnnnnnnnnnnnnnnnnnnnnnnnnnngtcgtctttgg 1715
           ||||||||||||||||||||||                            ||||||||||
Db      72 CTTTGGTTTGAGAGCCATGGCAAAAAAAAAAAAAAGAAAAAAAAGATCAAGTCGTCTTTGG 131

Qy    1716 tgagccagtaaggtgaaagcttgctgactgtccaaggcacaagagaaaattgaggaattg 1775
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 TGAGCCAGTAAGGTGAAAGCTTGCTGACTGTCCAAGGCACAAGAGAAAATTGAGGAATTG 191

Qy    1776 aaatgcaacctgagtatcaaactaaatattctaatcaaaggtaggtactgttaggtggaa 1835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 AAATGCAACCTGAGTATCAAACTAAATATTCTAATCAAAGGTAGGTACTGTTAGGTGGAA 251

Qy    1836 ttctatcagcaggcaactgcaaatgagaagaagatagaaggacgcccgtcgggactttgg 1895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     252 TTCTATCAGCAGGCAACTGCAAATGAGAAGAAGATAGAAGGACGCCCGTCGGGACTTTGG 311

Qy    1896 agggcagtgttatttttcccaaagaaagacggccaagggcagaggcatggattctttgcag 1955
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     312 AGGGCATTGTTATTTTCCCAAAGAAAGACGGCCAAGGGCAGAGGCATGGATTCTTTGCAG 371

Qy    1956 agcacttccttttggtttttcagtactgtttcatagacagtgggctcacatgttcctgat 2015
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 AGCACTTCCTTTTGGTTNTTCAGTACTGTTTCATAGACAGTGGGCTCACATGTTCCTGAT 431

Qy    2016 agtgctgcagttgcttagaaagcatcccagtt-attgcagtaattagaacttctggaata 2074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 AGTGCTGCAGTTGCTTAGAAAGCATCCCAGTTAATTGCAGTAATTAGAACTTCTGGAATA 491

Qy    2075 tgctagggcagaagtatgtcaagtatgtcacatgaagaaaatgtgaaattcaagagtaat 2134
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 TGCTGGGGCAGAAGTATGTCAAGTATGTCACATGAAGAAAATGTGAAATTCAAGAGTAAT 551

Qy    2135 ccacacgtgagaaact 2150
           ||||||||||||||||
Db     552 CCACACGTGAGAAACT 567


RESULT 12
AU144115/c
LOCUS       AU144115      571 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU144115 HEMBA1 Homo sapiens cDNA clone HEMBA1000916 3', mRNA
            sequence.
ACCESSION   AU144115
VERSION     AU144115.1  GI:11005636
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
            ,S. and Isogai,T.).
  TITLE     HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
            Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
            ,Y., Nagai,T., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1000916"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      185 a    115 c     89 g    172 t     10 others
ORIGIN


  Query Match             21.4%;  Score 498.8;  DB 107;  Length 571;
  Best Local Similarity   94.8%;  Pred. No. 6e-91;
  Matches  542;  Conservative    0;  Mismatches   27;  Indels    3;  Gaps    3;

Qy    1108 catattgagatttttttttaatcaacttttaagatagtgatgtt-ctgttctaaactgttc 1166
           || |||||  ||||||| |  || |||||||| | || || ||| | ||| ||| |||| |
Db     571 CAAATTGAATTTTTTTTTAANCCACCTTTTAANAAAGNGANGTTCCGGTTNTAACCTGTNC 512

Qy    1167 tgttttagtgaaggtagatttttataaaacaagcatggggattcttttctaaggtaatat 1226
           |||| ||| ||||||| |||| ||||| ||||||||||||||||||||||||||||||||
Db     511 NGTTTAAGTAAAGGTAGTTTTTAATAAACCAAGCATGGGGATTCTTTTCTAAGGTAATAT 452

Qy    1227 taatgagaagggaaaaaagtatctttaa-cagctctttgttgaagcctgtggtagcacat 1285
           ||||||||||||||||||   |||||| |||||||||||| |||| ||||| ||| |||
Db     451 TAATGAGAAGGGAAAAAAGATCCTTTAACCAGCTCTTTGTTNAAGCNTGTGG-AGCNCAT 393

Qy    1286 tatgtttataattgcacatgtgcacataatctattatgatccaatgcaaatacagctcca 1345
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 TATGTTTATAATTGCACATGTGCACATAATCTATTATGATCCAATGCAAATACAGCTCCA 333

Qy    1346 aaaatattaaatgtatatatatttttaaaatgcctgaggaaatacatttttcttaataaac 1405
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 AAAATATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACATTTTTCTTAATAAAC 273

Qy    1406 tgaagagtctcagtatggctattaaaataattattagcctcctgttgtgtggctgcaaaa 1465
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 TGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTTGTGTGGCTGCAAAA 213

Qy    1466 catcacaaagtgaccggtcttgagacctgtgaactgctgccctgtttagtaaataaaatt 1525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 CATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTTTAGTAAATAAAATT 153

Qy    1526 aatgcatttctagaggggggaatatctgccatccagtggtggaaatgtggagtaaagaagc 1585
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152 AATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATGTGGAGTAAAGAAGC 93

Qy    1586 tggtggtctgcttctgtgctgtatgccagccttttgccttaagttgagaggaggtcaact 1645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92 TGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTGCCTTAAGTTGAGAGGAGGTCAACT 33

Qy    1646 ttagctactgtctttggtttgagagccatggc 1677
           ||||||||||||||||||||||||||||||||
Db      32 TTAGCTACTGTCTTTGGTTTGAGAGCCATGGC 1


RESULT 13
AI380343/c
LOCUS       AI380343      478 bp    mRNA            EST       30-MAR-1999
DEFINITION  tf94h10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2106979 3',
            mRNA sequence.
ACCESSION   AI380343
VERSION     AI380343.1  GI:4190196
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 866   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 451.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2106979"
                      /clone_lib="NCI_CGAP_CLL1"
                      /tissue_type="B-cell, chronic lymphotic leukemia"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library is normalized, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT      155 a    103 c     75 g    145 t
ORIGIN


  Query Match             20.5%;  Score 476;  DB 19;  Length 478;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-86;
  Matches  476;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1202 tggggattcttttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctct 1261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 TGGGGATTCTTTTCTAAGGTAATATTAATGAGAAGGGAAAAAAGTATCTTTAACAGCTCT 419

Qy   1262 ttgttgaagcctgtggtagcacattatgtttataattgcacatgtgcacataatctatta 1321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 TTGTTGAAGCCTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTA 359

Qy   1322 tgatccaatgcaaatacagctccaaaaaatattaaatgtatatatattttaaaatgcctga 1381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TGATCCAATGCAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGA 299

Qy   1382 ggaaatacatttttcttaataaactgaagagtctcagtatggctattaaaataattatta 1441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 GGAAATACATTTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTA 239

Qy   1442 gcctcctgttgtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactg 1501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 GCCTCCTGTTGTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTG 179

Qy   1502 ctgccctgtttagtaaataaaattaatgcatttctagaggggggaatatctgccatccagt 1561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 CTGCCCTGTTTAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGT 119

Qy   1562 ggtggaaatgtggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttg 1621
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 GGTGGAAATGTGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTGCTGTATGCCAGCCTTTTG 59

Qy   1622 ccttaagttgagaggaggtcaactttagctactgtctttggtttgagagccatggc 1677
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     58 CCTTAAGTTGAGAGGAGGTCAACTTTAGCTACTGTCTTTGGTTTGAGAGCCATGGC 3


RESULT  14
AI927871/c
```

```
LOCUS       AI927871      467 bp    mRNA            EST       08-MAR-2000
DEFINITION  wo66c11.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2460308 3',
            mRNA sequence.
ACCESSION   AI927871
VERSION     AI927871.1  GI:5663835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 897   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 466.
FEATURES              Location/Qualifiers
     source           1. .467
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2460308"
                      /clone_lib="NCI_CGAP_Pr22"
                      /sex="male"
                      /tissue_type="normal prostate"
                      /lab_host="DH10B"
                      /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; 1st strand cDNA was prepared
                      from normal prostate bulk tissue, and was then primed with
                      a Not I - oligo(dT) primer. Double-stranded cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector. Library is normalized, and was
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      150 a     98 c     78 g    141 t
ORIGIN


  Query Match             19.7%;  Score 458;  DB 103;  Length 467;
  Best Local Similarity   98.9%;  Pred. No. 1.1e-82;
  Matches  461;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy   1212 tttctaaggtaatattaatgagaagggaaaaaagtatctttaacagctctttgttgaagc 1271
           ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    467 TTTCTAAGGTAATATTAATGAGAAGGGAATAAAGTATCTTTAACAGCTCTTTGTTGAAGC 408

Qy   1272 ctgtggtagcacattatgtttataattgcacatgtgcacataatctattatgatccaatg 1331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 CTGTGGTAGCACATTATGTTTATAATTGCACATGTGCACATAATCTATTATGATCCAATG 348

Qy   1332 caaatacagctccaaaaatattaaatgtatatatattttaaaatgcctgaggaaatacat 1391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    347 CAAATACAGCTCCAAAAATATTAAATGTATATATATTTTAAAATGCCTGAGGAAATACAT 288

Qy   1392 ttttcttaataaactgaagagtctcagtatggctattaaaataattattagcctcctgtt 1451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 TTTTCTTAATAAACTGAAGAGTCTCAGTATGGCTATTAAAATAATTATTAGCCTCCTGTT 228

Qy   1452 gtgtggctgcaaaacatcacaaagtgaccggtcttgagacctgtgaactgctgccctgtt 1511
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    227 GTGTGGCTGCAAAACATCACAAAGTGACCGGTCTTGAGACCTGTGAACTGCTGCCCTGTT 168
```

```
Qy   1512 tagtaaataaaattaatgcatttctagaggggaatatctgccatccagtggtggaaatg 1571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    167 TAGTAAATAAAATTAATGCATTTCTAGAGGGGGAATATCTGCCATCCAGTGGTGGAAATG 108

Qy   1572 tggagtaaagaagctggtggtctgcttctgtgctgtatgccagccttttgccttaagttg 1631
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    107 TGGAGTAAAGAAGCTGGTGGTCTGCTTCTGTCCTGTATGCCAGCCTTTTGCCTTAAGTTG 48

Qy   1632 agaggaggtcaactttagctactgtctttggtttgagagccatggc 1677
          ||||||||||||||| ||| ||||||||||||||||||||||||||
Db     47 AGAGGAGGTCAACTTCAGCCCCTGTCTTTGGTTTGAGAGCCATGGC 2


RESULT 15
BF185507
LOCUS       BF185507      897 bp    mRNA            EST       31-OCT-2000
DEFINITION  601814588F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4045293 5',
            mRNA sequence.
ACCESSION   BF185507
VERSION     BF185507.1  GI:11063966
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM870  row: o  column: 22
            High quality sequence stop: 538.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4045293"
                     /clone_lib="NIH_MGC_55"
                     /tissue_type="from acute myelogenous leukemia"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  14/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      285 a    130 c    174 g    305 t      3 others
ORIGIN


  Query Match             19.5%;  Score 454.2;  DB 145;  Length 897;
  Best Local Similarity   92.7%;  Pred. No. 6.5e-82;
  Matches  575;  Conservative    0;  Mismatches   34;  Indels   11;  Gaps    9;

Qy      70 attcccccaaacaacaatatttaattttcttagtaaaaagaataacagaatgcatcgtgg 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1 ATTCCCCCAAACAACAATATTTAATTTTCTTAGTAAAAAGAATAACAGAATGCATC-TGG 59

Qy     130 caatccttaagcaacattatctatgtggactgcttaaatcagcaaaacaccagaagtttg 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60 CAATCCTTAAGCAACATTATCTATGTGGACTGCTTAAATCAGCAAAACACCAGAAGTTTG 119

Qy     190 gttaacttgggcaatatgacaagtattacttttgggcaaaactactcattaagcaattt 249
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     120 GTTAACTTGGGCAATATGACAAGTATTAC-TTTTGGGCAAAACTACTCATTAAGCAATTC 178

Qy     250 ctctagtgtgtcggacacaaataggttctttatttttggcatgtatgccttttttattttc 309
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     179 CTCTAGTGTGTCGGACACAAATAGGTTCTTTA-TTTTGGCATGTATGCCTTTTTATTTTC 237

Qy     310 attcaatttttttttttttctcagacagacatagtagtaacgactagcattggaaaatataca 369
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     238 ATTCAATTTTTTTTTTTTCTCAGACAGACATAGTAGTAACGACTAGCATGGGAAAATACA 297

Qy     370 tatcactattcttggaatatttatggtcagtctacttttttagtagaatatttttggatag 429
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     298 TATCACTATTCTTGGAATATTTATGGTCAGTCTACTTTTAAGTAGAATATTTTGGGATAG 357

Qy     430 cgttgacacgatagatcttattccatacttctttatt-attgataatttttattttcattt 488
           ||| |||||||||||||||||||||||||||||||| || ||||||| |  |||   ||
Db     358 CGTGGACACGATAGATCTTATTCCATACTTCTTTATTAATGGATAATTGTTATTTGCATT 417

Qy     489 tttgctttcattattatacatattttggtggagaagaggttgggctttttttgaaagagac 548
           ||||||| |||||||||||||||  | ||||||||||||||| ||| ||||||||||
Db     418 TTTGCTTGCATTATTATACATATTGNGTGGGAGAAGAGGTTGGGC-TTTCTGAAAGAGAC 476

Qy     549 aaaaatttattataacactaaacactccttttttgacatattaaagcctttattccatct 608
           |||||||||||||| |||||||||||||||||||||| |||||| ||   |  | |||||
Db     477 -AAAATTTATTATAGCACTAAACACTCCTTTTTTGA-ATATTAAGGC---TTATTCATCT 531

Qy     609 ctcaagatatattataaaatttatttttttaatttaag-atttctgaattattttatctt 667
           |||||||||| |  ||| |||| |||||| |||| ||| ||||||||||||||| |||||
Db     532 CTCAAGATATTTATTAAGATTTCTTTTTTCAATTGAAGAATTTCTGAATTATTTAATCTT 591

Qy     668 aaattgtgattttaaacgag 687
           ||||||||||||||   ||
Db     592 AAATTGTGATTTTAACGAAG 611


Search completed: April 30, 2001, 12:17:41
Job time: 12530 sec
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:08:52 ; Search time 22667.1 Seconds
(without alignments)
173.596 Million cell updates/sec

Title: US-09-721-183-4
Perfect score: 639
Sequence: 1 ccagaaccgagtttaggtcc.....cacgaatggacggaggggat 639

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		501.8	78.5	168927	75	AC073204	AC073204 Homo sapi
2		350.4	54.8	2360	35	AF193056	Af193056 Homo sapi
3	c	133.2	20.8	168927	75	AC073204	AC073204 Homo sapi
4		55.8	8.7	72883	68	AC023432	AC023432 Homo sapi
5	c	45.2	7.1	7218	10	I66494	I66494 Sequence 14
6	c	42.2	6.6	140921	78	AC089987	AC089987 Homo sapi
7	c	42	6.6	53522	93	HUMPKD1GEN	L39891 Homo sapien
8	c	42	6.6	53526	9	AR097326	AR097326 Sequence
9	c	42	6.6	53577	9	AR097325	AR097325 Sequence
10		41.2	6.4	169713	64	AC015919	AC015919 Homo sapi
11		41.2	6.4	170758	85	AC004965	AC004965 Homo sapi

c	12	41	6.4	143255	64	AC015996
c	13	40.8	6.4	159002	86	AC006013
	14	40.4	6.3	42375	89	AL133273
c	15	40.2	6.3	162405	85	AC002352
c	16	40.2	6.3	191422	67	AC022318
c	17	40.2	6.3	206589	87	AC009483
c	18	40.2	6.3	243085	60	AC008440
	19	39.8	6.2	754	53	CNS072NG
	20	39.8	6.2	7218	10	I66494
	21	39.8	6.2	162473	86	AC005803
c	22	39.8	6.2	172804	76	AC079171
	23	39.4	6.2	62043	89	AL161893
c	24	39.2	6.1	85952	92	HSA373N24
	25	39.2	6.1	105153	60	AC002979
	26	39.2	6.1	161745	77	AC084358
c	27	39	6.1	157176	77	AC084716
c	28	38.8	6.1	67780	78	AC087275
	29	38.8	6.1	134640	88	AC027130
	30	38.6	6.0	54567	77	AC083934
c	31	38.6	6.0	182414	73	AC055877
c	32	38.6	6.0	239395	73	AC063918
	33	38.2	6.0	895	53	CNS06LLW
	34	38.2	6.0	1662	92	HSCYSTC1
	35	38.2	6.0	7292	92	HSCST3G
c	36	38.2	6.0	141079	92	HSA218C14
c	37	38	5.9	83495	4	AC087076
c	38	38	5.9	180971	88	AC079310
c	39	37.8	5.9	245265	63	AC013372
c	40	37.6	5.9	169950	77	AC084049
c	41	37.6	5.9	195566	61	AC009590
c	42	37.4	5.9	102579	85	AC002375
c	43	37.4	5.9	116379	82	AP000783
c	44	37.4	5.9	140176	64	AC016411
c	45	37.4	5.9	150384	82	AP000841

AC015996 Homo sapi
AC006013 Homo sapi
AL133273 Human DNA
AC002352 Homo sapi
AC022318 Homo sapi
AC009483 Homo sapi
AC008440 Homo sapi
AL426434 clone BA0
I66494 Sequence 14
AC005803 Homo sapi
AC079171 Homo sapi
AL161893 Human DNA
AL121932 Human DNA
AC002979 Homo sapi
AC084358 Homo sapi
AC084716 Homo sapi
AC087275 Homo sapi
AC027130 Homo sapi
AC083934 Homo sapi
AC055877 Homo sapi
AC063918 Homo sapi
AL404346 T7 end of
X61681 Homo sapien
X52255 Human CST3
AL121894 Human DNA
AC087076 Caenorhab
AC079310 Homo sapi
AC013372 Homo sapi
AC084049 Homo sapi
AC009590 Homo sapi
AC002375 Homo sapi
AP000783 Homo sapi
AC016411 Homo sapi
AP000841 Homo sapi

ALIGNMENTS

RESULT 1
AC073204
LOCUS AC073204 168927 bp DNA HTG 12-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-316M1, WORKING DRAFT SEQUENCE, 46 unordered pieces.
ACCESSION AC073204
VERSION AC073204.3 GI:9053126
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168927)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 168927)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 12, 2000 this sequence version replaced gi:8571804.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0316M01
---------- Summary Statistics ----------
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 135538 bases at least Q40
Consensus quality: 144366 bases at least Q30
Consensus quality: 148947 bases at least Q20
Insert size: 193000; agarose-fp
Insert size: 164427; sum-of-contigs
Quality coverage: 3.10 in Q20 bases; agarose-fp
Quality coverage: 3.89 in Q20 bases; sum-of-contigs

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1741: contig of 1741 bp in length
*     1742     1841: gap of unknown length
*     1842     3024: contig of 1183 bp in length
*     3025     3124: gap of unknown length
*     3125     5533: contig of 2409 bp in length
*     5534     5633: gap of unknown length
*     5634     7851: contig of 2218 bp in length
*     7852     7951: gap of unknown length
*     7952     9380: contig of 1429 bp in length
*     9381     9480: gap of unknown length
*     9481    11067: contig of 1587 bp in length
*    11068    11167: gap of unknown length
*    11168    12450: contig of 1283 bp in length
*    12451    12550: gap of unknown length
*    12551    14241: contig of 1691 bp in length
*    14242    14341: gap of unknown length
*    14342    16039: contig of 1698 bp in length
*    16040    16139: gap of unknown length
*    16140    17399: contig of 1260 bp in length
*    17400    17499: gap of unknown length
*    17500    19843: contig of 2344 bp in length
*    19844    19943: gap of unknown length
*    19944    22794: contig of 2851 bp in length
*    22795    22894: gap of unknown length
*    22895    24627: contig of 1733 bp in length
*    24628    24727: gap of unknown length
*    24728    27062: contig of 2335 bp in length
*    27063    27162: gap of unknown length
*    27163    28786: contig of 1624 bp in length
*    28787    28886: gap of unknown length
*    28887    31043: contig of 2157 bp in length
*    31044    31143: gap of unknown length
*    31144    32758: contig of 1615 bp in length
*    32759    32858: gap of unknown length
*    32859    35681: contig of 2823 bp in length
*    35682    35781: gap of unknown length
*    35782    38247: contig of 2466 bp in length
*    38248    38347: gap of unknown length
*    38348    40598: contig of 2251 bp in length
*    40599    40698: gap of unknown length
*    40699    42593: contig of 1895 bp in length
*    42594    42693: gap of unknown length
*    42694    46042: contig of 3349 bp in length
*    46043    46142: gap of unknown length
*    46143    48549: contig of 2407 bp in length
*    48550    48649: gap of unknown length
*    48650    51055: contig of 2406 bp in length
*    51056    51155: gap of unknown length
*    51156    55216: contig of 4061 bp in length
*    55217    55316: gap of unknown length
*    55317    57842: contig of 2526 bp in length
*    57843    57942: gap of unknown length
*    57943    60798: contig of 2856 bp in length
*    60799    60898: gap of unknown length
*    60899    64058: contig of 3160 bp in length
```

```
            *    64059    64158: gap of unknown length
            *    64159    67583: contig of 3425 bp in length
            *    67584    67683: gap of unknown length
            *    67684    70781: contig of 3098 bp in length
            *    70782    70881: gap of unknown length
            *    70882    74607: contig of 3726 bp in length
            *    74608    74707: gap of unknown length
            *    74708    79568: contig of 4861 bp in length
            *    79569    79668: gap of unknown length
            *    79669    82429: contig of 2761 bp in length
            *    82430    82529: gap of unknown length
            *    82530    86716: contig of 4187 bp in length
            *    86717    86816: gap of unknown length
            *    86817    91042: contig of 4226 bp in length
            *    91043    91142: gap of unknown length
            *    91143    94644: contig of 3502 bp in length
            *    94645    94744: gap of unknown length
            *    94745   100042: contig of 5298 bp in length
            *   100043   100142: gap of unknown length
            *   100143   105920: contig of 5778 bp in length
            *   105921   106020: gap of unknown length
            *   106021   110097: contig of 4077 bp in length
            *   110098   110197: gap of unknown length
            *   110198   114876: contig of 4679 bp in length
            *   114877   114976: gap of unknown length
            *   114977   121638: contig of 6662 bp in length
            *   121639   121738: gap of unknown length
            *   121739   126383: contig of 4645 bp in length
            *   126384   126483: gap of unknown length
            *   126484   130682: contig of 4199 bp in length
            *   130683   130782: gap of unknown length
            *   130783   138712: contig of 7930 bp in length
            *   138713   138812: gap of unknown length
            *   138813   148694: contig of 9882 bp in length
            *   148695   148794: gap of unknown length
            *   148795   168927: contig of 20133 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168927
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-316M1"
BASE COUNT    43817 a  38627 c  38053 g  43921 t   4509 others
ORIGIN
```

Query Match 78.5%; Score 501.8; DB 75; Length 168927;
Best Local Similarity 99.6%; Pred. No. 1.1e-136;
Matches 503; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 ccagaaccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114273 CCAGAACCGAGTTTAGGTCCAGGTTCTCGTTCTGGCAAATCTTTCTCCTTACCTTCTTCC 114332

Qy      61 tccacccctccacctatgccatgttttcccttagccactccccagctcggtggaggaaag 120
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  114333 TCCACCCCTCCACTTATGCCATGTTTTCCCTTAGCCACTCCCCAGCTCGGTGGAGGAAAG 114392

Qy     121 gcaggcctaactaggtaccgtcttcccgactttgctcaatgatagctgggtgggtctagc 180
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  114393 GCAGGCCTAACTAGGTAGCGTCTTCCCGACTTTGCTCAATGATAGCTGGGTGGGTCTAGC 114452

Qy     181 tgggttccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114453 TGGGTTCCAGCCACTTGTAATGTGGGACATCTCTCACCCCAACTTTGTAGGTGGAGCAAC 114512

Qy     241 tgctacagaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttaa 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114513 TGCTACAGAGGTAAATATGATTAACTTTACATTCCATCTTTCGTCTGCTCCCAAACTTAA 114572

Qy     301 cagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtataggaggca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114573 CAGCAGGTAATCTGCTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATCTGTATAGGAGGCA 114632

Qy     361 agagatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgttcttca 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114633 AGAGATCTGAGTCCTTTTGAAGGCCTATCCTCTGCTCTGTATCTCAATTACTGTTCTTCA 114692

Qy     421 tttcaattattcttacctactattcagttcccttgatcttttcttcttgggggctgtctt 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114693 TTTCAATTATTCTTACCTACTATTCAGTTCCCTTGATCTTTTCTTCTTGGGGGCTGTCTT 114752

Qy     481 agggtcagggagattgcagaagcac 505
           |||||||||||||||||||||||||
Db  114753 AGGGTCAGGGAGATTGCAGAAGCAC 114777
```

```
RESULT   2
AF193056
ID   AF193056   standard; RNA; HUM; 2360 BP.
XX
AC   AF193056;
XX
SV   AF193056.1
XX
DT   13-OCT-2000 (Rel. 65, Created)
DT   13-OCT-2000 (Rel. 65, Last updated, Version 1)
XX
DE   Homo sapiens PP753 mRNA, complete cds.
XX
KW   .
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-2360
RA   Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P., Qin W.X.,
RA   Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y., Yu J., Han L.H.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell growth";
RL   Unpublished.
XX
RN   [2]
RP   1-2360
RA   Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P., Qin W.X.,
RA   Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y., Yu J., Han L.H.;
RT   ;
RL   Submitted (08-OCT-1999) to the EMBL/GenBank/DDBJ databases.
RL   National Laboratory For Oncogenes & Related Genes, Shanghai Cancer
RL   Institute, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .2360
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT   CDS             618. .1445
FT                   /codon_start=1
FT                   /gene="PP753"
FT                   /product="unknown"
FT                   /protein_id="AAG22484.1"
FT                   /translation="MRKRLSAPELRLSLTKGPGNDGASPTQSAPSSPDGSSDLEIDELE
FT                   TPSDSEQLDSGHEFEWEDELPRAEGLGTSETAERLGRGCMWDVTGEDGHHWRVFRMGPR
FT                   EQRVDMTVIEPYKKVLSHGGYHGDGLNAVILFASCYLPRSSIPNYTYVMEHLFRYMVGT
FT                   LELLVAENYLLVHLSGGTSRAQVPPLSWIRQCYRTLDRRLRKNLRALVVVHATWYVKAF
FT                   LALLRPFISSKFTRKIRFLDSLGELAQLISLDQVHIPEAVRQLDRDLHGSGGT"
XX
SQ   Sequence 2360 BP; 569 A; 562 C; 621 G; 608 T; 0 other;
```

Query Match 54.8%; Score 350.4; DB 35; Length 2360;
Best Local Similarity 97.0%; Pred. No. 3.2e-92;
Matches 357; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy     228 taggtggagcaactgctacagaggtaaatatgattaactttacattccatctttcgtctg 287
```

```
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 TGGGTGGAGCAACTGCTACAGAGGTAAATATGATTAACTTTACATTCCATCTTTCGTCTG 217

Qy    288 ctcccaaacttaacagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatc 347
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 CTCCCAAACTTAACAGCAGGTAATCTGCTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATC 277

Qy    348 tgtataggaggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaa 407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 TGTATAGGAGGCAAGAGATCTGAGTCCTTTTGAAGGCCTATCCTCTGCTCTGTATCTCAA 337

Qy    408 ttactgttcttcatttcaattattcttacctactattcagttcccttgatcttttcttct 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    338 TTACTGTTCTTCATTTCAATTATTCTTACCTACTATTCAGTTCCCTTGATCTTTTCTTCT 397

Qy    468 tgggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgag 527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    398 TGGGGGCTGTCTTAGGGTCAGGGAGATTGCAGAAGCACCAGAACTAGGAGCAGCCCTGAG 457

Qy    528 acatggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacg 587
          |||||||||||||||||||||||||||||||||||||||||||||||||| |     |
Db    458 ACATGGGGAGTTGGAGCTGAAGGAGGAATGGCAGGATGAAGAATTCCCTAGATTGCTTCC 517

Qy    588 tgtgaggg 595
          || |  ||
Db    518 TGAGGAGG 525


RESULT   3
AC073204/c
LOCUS       AC073204   168927 bp    DNA             HTG       12-JUL-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-316M1, WORKING DRAFT SEQUENCE,
            46 unordered pieces.
ACCESSION   AC073204
VERSION     AC073204.3  GI:9053126
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168927)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168927)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 12, 2000 this sequence version replaced gi:8571804.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0316M01
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 135538 bases at least Q40
            Consensus quality: 144366 bases at least Q30
            Consensus quality: 148947 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 164427; sum-of-contigs
            Quality coverage: 3.10 in Q20 bases; agarose-fp
            Quality coverage: 3.89 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1741: contig of 1741 bp in length
            *     1742     1841: gap of unknown length
            *     1842     3024: contig of 1183 bp in length
            *     3025     3124: gap of unknown length
            *     3125     5533: contig of 2409 bp in length
            *     5534     5633: gap of unknown length
            *     5634     7851: contig of 2218 bp in length
            *     7852     7951: gap of unknown length
            *     7952     9380: contig of 1429 bp in length
            *     9381     9480: gap of unknown length
            *     9481    11067: contig of 1587 bp in length
            *    11068    11167: gap of unknown length
            *    11168    12450: contig of 1283 bp in length
            *    12451    12550: gap of unknown length
            *    12551    14241: contig of 1691 bp in length
            *    14242    14341: gap of unknown length
            *    14342    16039: contig of 1698 bp in length
            *    16040    16139: gap of unknown length
            *    16140    17399: contig of 1260 bp in length
            *    17400    17499: gap of unknown length
            *    17500    19843: contig of 2344 bp in length
            *    19844    19943: gap of unknown length
            *    19944    22794: contig of 2851 bp in length
            *    22795    22894: gap of unknown length
            *    22895    24627: contig of 1733 bp in length
            *    24628    24727: gap of unknown length
            *    24728    27062: contig of 2335 bp in length
            *    27063    27162: gap of unknown length
            *    27163    28786: contig of 1624 bp in length
            *    28787    28886: gap of unknown length
            *    28887    31043: contig of 2157 bp in length
            *    31044    31143: gap of unknown length
            *    31144    32758: contig of 1615 bp in length
            *    32759    32858: gap of unknown length
            *    32859    35681: contig of 2823 bp in length
            *    35682    35781: gap of unknown length
            *    35782    38247: contig of 2466 bp in length
            *    38248    38347: gap of unknown length
            *    38348    40598: contig of 2251 bp in length
            *    40599    40698: gap of unknown length
            *    40699    42593: contig of 1895 bp in length
            *    42594    42693: gap of unknown length
            *    42694    46042: contig of 3349 bp in length
            *    46043    46142: gap of unknown length
            *    46143    48549: contig of 2407 bp in length
            *    48550    48649: gap of unknown length
            *    48650    51055: contig of 2406 bp in length
            *    51056    51155: gap of unknown length
            *    51156    55216: contig of 4061 bp in length
            *    55217    55316: gap of unknown length
            *    55317    57842: contig of 2526 bp in length
            *    57843    57942: gap of unknown length
            *    57943    60798: contig of 2856 bp in length
            *    60799    60898: gap of unknown length
            *    60899    64058: contig of 3160 bp in length
            *    64059    64158: gap of unknown length
            *    64159    67583: contig of 3425 bp in length
            *    67584    67683: gap of unknown length
            *    67684    70781: contig of 3098 bp in length
            *    70782    70881: gap of unknown length
            *    70882    74607: contig of 3726 bp in length
            *    74608    74707: gap of unknown length
            *    74708    79568: contig of 4861 bp in length
            *    79569    79668: gap of unknown length
            *    79669    82429: contig of 2761 bp in length
            *    82430    82529: gap of unknown length
```

```
            *    82530    86716: contig of 4187 bp in length
            *    86717    86816: gap of unknown length
            *    86817    91042: contig of 4226 bp in length
            *    91043    91142: gap of unknown length
            *    91143    94644: contig of 3502 bp in length
            *    94645    94744: gap of unknown length
            *    94745   100042: contig of 5298 bp in length
            *   100043   100142: gap of unknown length
            *   100143   105920: contig of 5778 bp in length
            *   105921   106020: gap of unknown length
            *   106021   110097: contig of 4077 bp in length
            *   110098   110197: gap of unknown length
            *   110198   114876: contig of 4679 bp in length
            *   114877   114976: gap of unknown length
            *   114977   121638: contig of 6662 bp in length
            *   121639   121738: gap of unknown length
            *   121739   126383: contig of 4645 bp in length
            *   126384   126483: gap of unknown length
            *   126484   130682: contig of 4199 bp in length
            *   130683   130782: gap of unknown length
            *   130783   138712: contig of 7930 bp in length
            *   138713   138812: gap of unknown length
            *   138813   148694: contig of 9882 bp in length
            *   148695   148794: gap of unknown length
            *   148795   168927: contig of 20133 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168927
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-316M1"
BASE COUNT    43817 a  38627 c  38053 g  43921 t   4509 others
ORIGIN


  Query Match             20.8%;  Score 133.2;  DB 75;  Length 168927;
  Best Local Similarity   58.4%;  Pred. No. 4.1e-28;
  Matches  146;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy     390 ctctgctctgtatctcaattactgttcttcatttcaattattcttacctactattcagtt 449
           || ||
Db   57947 CTTTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57888

Qy     450 cccttgatcttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccaga 509
                                                     | ||||||||||||||
Db   57887 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTCAGAAGCACCAGA 57828

Qy     510 actaggagcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaaga 569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57827 ACTAGGAGCAGCCCTGAGACATGGGGAGTTGGAGCTGAAGGAGGAATGGCAGGATGAAGA 57768

Qy     570 attccctaggtgaggacgtgtgagggtggctgggagaagggaggggtggtcacgaatgga 629
           |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db   57767 ATTCCCTAGGTGAGGACTTGTGAGGGTGGCTGGGAGAAGGGA-GGGTGGTCACGAATGGA 57709

Qy     630 cggaggggat 639
           ||||||||||
Db   57708 CGGAGGGGAT 57699


RESULT   4
AC023432
LOCUS       AC023432    72883 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-115I7, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023432
VERSION     AC023432.2  GI:9112064
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 72883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-115I7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6970641.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1877
                Center clone name: 115_I_7
            --------------
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      730: contig of 730 bp in length
            *      731 830: gap of      100 bp
            *      831     1580: contig of 750 bp in length
            *     1581 1680: gap of      100 bp
            *     1681     2452: contig of 772 bp in length
            *     2453 2552: gap of      100 bp
            *     2553     3323: contig of 771 bp in length
            *     3324 3423: gap of      100 bp
            *     3424     4159: contig of 736 bp in length
            *     4160 4259: gap of      100 bp
            *     4260     4999: contig of 740 bp in length
            *     5000 5099: gap of      100 bp
            *     5100     5860: contig of 761 bp in length
            *     5861 5960: gap of      100 bp
            *     5961     6726: contig of 766 bp in length
            *     6727 6826: gap of      100 bp
            *     6827     7586: contig of 760 bp in length
            *     7587 7686: gap of      100 bp
            *     7687     8449: contig of 763 bp in length
            *     8450 8549: gap of      100 bp
            *     8550     9311: contig of 762 bp in length
            *     9312 9411: gap of      100 bp
```

```
*     9412    10155: contig of 744 bp in length
*    10156 10255: gap of      100 bp
*    10256    11019: contig of 764 bp in length
*    11020 11119: gap of      100 bp
*    11120    11895: contig of 776 bp in length
*    11896 11995: gap of      100 bp
*    11996    12778: contig of 783 bp in length
*    12779 12878: gap of      100 bp
*    12879    13629: contig of 751 bp in length
*    13630 13729: gap of      100 bp
*    13730    14493: contig of 764 bp in length
*    14494 14593: gap of      100 bp
*    14594    15341: contig of 748 bp in length
*    15342 15441: gap of      100 bp
*    15442    16206: contig of 765 bp in length
*    16207 16306: gap of      100 bp
*    16307    17072: contig of 766 bp in length
*    17073 17172: gap of      100 bp
*    17173    17931: contig of 759 bp in length
*    17932 18031: gap of      100 bp
*    18032    18780: contig of 749 bp in length
*    18781 18880: gap of      100 bp
*    18881    19652: contig of 772 bp in length
*    19653 19752: gap of      100 bp
*    19753    20428: contig of 676 bp in length
*    20429 20528: gap of      100 bp
*    20529    21305: contig of 777 bp in length
*    21306 21405: gap of      100 bp
*    21406    22161: contig of 756 bp in length
*    22162 22261: gap of      100 bp
*    22262    23030: contig of 769 bp in length
*    23031 23130: gap of      100 bp
*    23131    23883: contig of 753 bp in length
*    23884 23983: gap of      100 bp
*    23984    24744: contig of 761 bp in length
*    24745 24844: gap of      100 bp
*    24845    25599: contig of 755 bp in length
*    25600 25699: gap of      100 bp
*    25700    26453: contig of 754 bp in length
*    26454 26553: gap of      100 bp
*    26554    27308: contig of 755 bp in length
*    27309 27408: gap of      100 bp
*    27409    28177: contig of 769 bp in length
*    28178 28277: gap of      100 bp
*    28278    29065: contig of 788 bp in length
*    29066 29165: gap of      100 bp
*    29166    29934: contig of 769 bp in length
*    29935 30034: gap of      100 bp
*    30035    30799: contig of 765 bp in length
*    30800 30899: gap of      100 bp
*    30900    31661: contig of 762 bp in length
*    31662 31761: gap of      100 bp
*    31762    32529: contig of 768 bp in length
*    32530 32629: gap of      100 bp
*    32630    33403: contig of 774 bp in length
*    33404 33503: gap of      100 bp
*    33504    34267: contig of 764 bp in length
*    34268 34367: gap of      100 bp
*    34368    35116: contig of 749 bp in length
*    35117 35216: gap of      100 bp
*    35217    35960: contig of 744 bp in length
*    35961 36060: gap of      100 bp
*    36061    36830: contig of 770 bp in length
*    36831 36930: gap of      100 bp
*    36931    37703: contig of 773 bp in length
*    37704 37803: gap of      100 bp
*    37804    38572: contig of 769 bp in length
*    38573 38672: gap of      100 bp
*    38673    39404: contig of 732 bp in length
*    39405 39504: gap of      100 bp
*    39505    40250: contig of 746 bp in length
*    40251 40350: gap of      100 bp
*    40351    41106: contig of 756 bp in length
*    41107 41206: gap of      100 bp
*    41207    41968: contig of 762 bp in length
*    41969 42068: gap of      100 bp
*    42069    42837: contig of 769 bp in length
*    42838 42937: gap of      100 bp
*    42938    43693: contig of 756 bp in length
*    43694 43793: gap of      100 bp
*    43794    44560: contig of 767 bp in length
*    44561 44660: gap of      100 bp
*    44661    45408: contig of 748 bp in length
*    45409 45508: gap of      100 bp
*    45509    46261: contig of 753 bp in length
*    46262 46361: gap of      100 bp
*    46362    47128: contig of 767 bp in length
*    47129 47228: gap of      100 bp
*    47229    48001: contig of 773 bp in length
*    48002 48101: gap of      100 bp
*    48102    48861: contig of 760 bp in length
*    48862 48961: gap of      100 bp
*    48962    49725: contig of 764 bp in length
*    49726 49825: gap of      100 bp
*    49826    50595: contig of 770 bp in length
*    50596 50695: gap of      100 bp
*    50696    51459: contig of 764 bp in length
*    51460 51559: gap of      100 bp
*    51560    52323: contig of 764 bp in length
*    52324 52423: gap of      100 bp
*    52424    53174: contig of 751 bp in length
*    53175 53274: gap of      100 bp
*    53275    54004: contig of 730 bp in length
*    54005 54104: gap of      100 bp
*    54105    54873: contig of 769 bp in length
*    54874 54973: gap of      100 bp
*    54974    55748: contig of 775 bp in length
*    55749 55848: gap of      100 bp
*    55849    56610: contig of 762 bp in length
*    56611 56710: gap of      100 bp
*    56711    57461: contig of 751 bp in length
*    57462 57561: gap of      100 bp
*    57562    58301: contig of 740 bp in length
*    58302 58401: gap of      100 bp
*    58402    59164: contig of 763 bp in length
*    59165 59264: gap of      100 bp

  Query Match              8.7%;  Score 55.8;  DB 68;  Length 72883;
  Best Local Similarity   35.8%;  Pred. No. 2.4e-05;
  Matches   57;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       1 ccagaaccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50540 CCAGAACCGAGTTTAGGTCCAGGTTCTCGTTCTGGCAAATCTTTCTCCTTACCTTANNNN 50599

Qy      61 tccacccctccacctatgccatgttttcccttagccactccccagctcggtggaggaaag 120

Db   50600 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 50659

Qy     121 gcaggcctaactaggtaccgtcttcccgactttgctcaa 159
                                                ||
Db   50660 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAA 50698


RESULT   5
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
```

```
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             7.1%;  Score 45.2;  DB 10;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 0.027;
  Matches    5;  Conservative  111;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     479 ttagggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagt 538
           || ||  |::::::::  : ::::: :  ::::  ::::: ::    :::: : ::::::
Db    1442 TTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1383

Qy     539 tggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtgg 598
            ::::  :::::::::: :: :::: :::::::      ::: ::::: : : ::::: ::
Db    1382 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1323

Qy     599 ctgggagaagggaggggtggtcacgaatggacggagggga 638
           ::::::::::::::: ::  : ::: ::: ::::::::
Db    1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1283


RESULT   6
AC089987/c
LOCUS       AC089987   140921 bp    DNA             HTG       12-FEB-2001
DEFINITION  Homo sapiens chromosome 12q clone RP11-722P11, WORKING DRAFT
            SEQUENCE, 45 unordered pieces.
ACCESSION   AC089987
VERSION     AC089987.3  GI:12745057
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140921)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140921)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 11, 2001 this sequence version replaced gi:12739552.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: HCVJ
                Center clone name: RP11-722P11
            ------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 130464 bases at least Q40
                Consensus quality: 139290 bases at least Q30
                Consensus quality: 143022 bases at least Q20
                Estimated insert size: 128307; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 45 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8789: contig of 8789 bp in length
            *     8790     8889: gap of unknown length
            *     8890    15186: contig of 6297 bp in length
            *    15187    15286: gap of unknown length
            *    15287    21703: contig of 6417 bp in length
            *    21704    21803: gap of unknown length
            *    21804    29293: contig of 7490 bp in length
            *    29294    29393: gap of unknown length
            *    29394    36986: contig of 7593 bp in length
            *    36987    37086: gap of unknown length
            *    37087    42749: contig of 5663 bp in length
            *    42750    42849: gap of unknown length
            *    42850    47700: contig of 4851 bp in length
            *    47701    47800: gap of unknown length
            *    47801    53388: contig of 5588 bp in length
            *    53389    53488: gap of unknown length
            *    53489    57678: contig of 4190 bp in length
            *    57679    57778: gap of unknown length
            *    57779    61697: contig of 3919 bp in length
            *    61698    61797: gap of unknown length
            *    61798    65970: contig of 4173 bp in length
            *    65971    66070: gap of unknown length
            *    66071    69288: contig of 3218 bp in length
            *    69289    69388: gap of unknown length
            *    69389    72476: contig of 3088 bp in length
            *    72477    72576: gap of unknown length
            *    72577    76245: contig of 3669 bp in length
            *    76246    76345: gap of unknown length
            *    76346    78992: contig of 2647 bp in length
            *    78993    79092: gap of unknown length
```

```
            *    79093    82091: contig of 2999 bp in length
            *    82092    82191: gap of unknown length
            *    82192    85068: contig of 2877 bp in length
            *    85069    85168: gap of unknown length
            *    85169    88153: contig of 2985 bp in length
            *    88154    88253: gap of unknown length
            *    88254    90751: contig of 2498 bp in length
            *    90752    90851: gap of unknown length
            *    90852    93038: contig of 2187 bp in length
            *    93039    93138: gap of unknown length
            *    93139    95078: contig of 1940 bp in length
            *    95079    95178: gap of unknown length
            *    95179    98237: contig of 3059 bp in length
            *    98238    98337: gap of unknown length
            *    98338   101202: contig of 2865 bp in length
            *   101203   101302: gap of unknown length
            *   101303   104017: contig of 2715 bp in length
            *   104018   104117: gap of unknown length
            *   104118   105988: contig of 1871 bp in length
            *   105989   106088: gap of unknown length
            *   106089   108376: contig of 2288 bp in length
            *   108377   108476: gap of unknown length
            *   108477   110968: contig of 2492 bp in length
            *   110969   111068: gap of unknown length
            *   111069   113117: contig of 2049 bp in length
            *   113118   113217: gap of unknown length
            *   113218   114685: contig of 1468 bp in length
            *   114686   114785: gap of unknown length
            *   114786   117181: contig of 2396 bp in length
            *   117182   117281: gap of unknown length
            *   117282   118973: contig of 1692 bp in length
            *   118974   119073: gap of unknown length
            *   119074   120761: contig of 1688 bp in length
            *   120762   120861: gap of unknown length
            *   120862   122703: contig of 1842 bp in length
            *   122704   122803: gap of unknown length
            *   122804   124053: contig of 1250 bp in length
            *   124054   124153: gap of unknown length
            *   124154   125515: contig of 1362 bp in length
            *   125516   125615: gap of unknown length
            *   125616   127736: contig of 2121 bp in length
            *   127737   127836: gap of unknown length
            *   127837   129233: contig of 1397 bp in length
            *   129234   129333: gap of unknown length
            *   129334   130892: contig of 1559 bp in length
            *   130893   130992: gap of unknown length
            *   130993   132449: contig of 1457 bp in length
            *   132450   132549: gap of unknown length
            *   132550   134257: contig of 1708 bp in length
            *   134258   134357: gap of unknown length
            *   134358   135517: contig of 1160 bp in length
            *   135518   135617: gap of unknown length
            *   135618   137184: contig of 1567 bp in length
            *   137185   137284: gap of unknown length
            *   137285   138362: contig of 1078 bp in length
            *   138363   138462: gap of unknown length
            *   138463   139817: contig of 1355 bp in length
            *   139818   139917: gap of unknown length
            *   139918   140921: contig of 1004 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140921
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-722P11"
BASE COUNT    39162 a  28247 c  27979 g  41018 t   4515 others
ORIGIN


  Query Match             6.6%;  Score 42.2;  DB 78;  Length 140921;
  Best Local Similarity   55.0%;  Pred. No. 0.26;
  Matches   83;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;
```

```
Qy     487 agggagattgcagaagcaccagaactaggagcagccctgagacatggggagttggagctg 546
           |||| ||| |  | |     || |  | | |  |     ||  ||||||  | ||||  |
Db 137175 AGGGGGATAGAGGGAAAGGAAGGAAAAAGGGAGGAGAGAAGGGATGGGGGATGGGAGGGG 137116

Qy     547 aaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtggctgggaga 606
           | ||||| | || || | || ||      |    ||||  | | |||||  |   ||  |
Db 137115 AGGGAGGGAAGGAAGAAGGATGATGGAAATGATGGAGGGTGGGAGAGGGGAGTAAGGGAA 137056

Qy     607 agggaggggtggtcacgaatggacggagggg 637
           ||||||| | | |   |   ||||| | |||
Db 137055 AGGGAGGAGGGATGGGGGGGGGACGTACGGG 137025


RESULT   7
HUMPKD1GEN/c
LOCUS       HUMPKD1GEN  53522 bp    DNA             PRI       19-JUN-1995
DEFINITION  Homo sapiens polycystic kidney disease-associated protein (PKD1)
            gene, complete cds.
ACCESSION   L39891
VERSION     L39891.1  GI:790818
KEYWORDS    polycystic kidney disease.
SOURCE      Homo sapiens (clone: cggg10.2, cDEB11) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ward,C.J., Peral,B., Hughes,J., Thomas,S., Gamble,V.,
            MacCarthy,A.B., Sloane-Stanley,J., Buckle,V.J., Kearney,L.,
            Higgs,D.R., Ratcliffe,P.J., Harris,P.C., Roelfsema,J.H.,
            Spruit,L.L., Saris,J.J., Dauwerse,H.G., Peters,D.J.M.,
            Breuning,M.H., Nellist,M., Brook-Carter,P.T., Maheshwar,M.M.,
            Cordeiro,I., Santos,H., Cabral,P., Sampson,J.R., Janssen,B.,
            Hesseling-Janssen,A.L.W., van den Ouweland,A.M.W., Eussen,B.,
            Verhoef,S., Lindhout,D. and Halley,D.J.J.
  TITLE     The polycystic kidney disease 1 gene encodes a 14 kb transcript and
            lies within a duplicated region on chromosome 16. The European
            Polycystic Kidney Disease Consortium [published erratum appears in
            Cell 1994 Aug 26;78(4):following 724]
  JOURNAL   Cell 77 (6), 881-894 (1994)
  MEDLINE   94273192
REFERENCE   2  (bases 1 to 53522)
  AUTHORS   Burn,T.C., Connors,T.D., Dackowski,W.R., Petry,L.R., Van Raay,T.J.,
            Millholland,J., Venet,M., Miller,G., Hakim,R.M., Doggett,N.A.,
            Landes,G.M., Klinger,K.W., Qian,F., Onuchic,L.F., Watnick,T. and
            Germino,G.G.
  TITLE     Analysis of the genomic sequence for the autosomal dominant
            polycystic kidney disease (PKD1) gene predicts the presence of a
            leucine-rich repeat. The American PKD1 Consortium (APKD1
            Consortium)
  JOURNAL   Hum. Mol. Genet. 4 (4), 575-582 (1995)
  MEDLINE   95359960
FEATURES             Location/Qualifiers
     source          1. .53522
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="cggg10.2, cDEB11"
                     /map="16p13.3"
     repeat_region   complement(236. .524)
                     /rpt_family="Alu"
     repeat_region   complement(704. .940)
                     /rpt_family="Alu"
     repeat_region   complement(998. .1287)
                     /rpt_family="Alu"
     repeat_region   complement(1483. .1771)
                     /rpt_family="Alu"
     repeat_region   complement(2081. .2352)
                     /rpt_family="Alu"
     repeat_region   complement(2423. .2666)
                     /rpt_family="Alu"
     misc_feature    3096. .4204
                     /note="CpG island; putative"
     exon            <3648. .3862
                     /gene="PKD1"
```

```
                /note="putative"
                /number=1
     gene       join(3648. .3862,15765. .15817,19866. .19974,20096. .20167,
                20436. .20605,21110. .21490,21609. .21792,22228. .22448,
                22637. .22752,23118. .23289,23656. .23815,24967. .25111,
                25989. .26120,26318. .26493,26808. .26941,27407. .27525,
                27628. .28683,29126. .29875,30419. .31024,31244. .31393,
                32328. .32471,32599. .32878,32972. .33185,33251. .33410,
                33848. .33952,36954. .37098,37701. .38330,38624. .38780,
                38961. .39213,39338. .39533,41028. .41198,41285. .41428,
                41523. .41733,41824. .41950,43610. .43726,43817. .43866,
                44091. .44275,44353. .44446,47384. .47502,47581. .47783,
                47856. .48050,48501. .48640,49002. .49114,49406. .49547,
                49688. .49813,49997. .50171,50419. .50709,50785. .50919,
                51003. .51308,51399. .51866)
                /gene="PKD1"
     CDS        join(3648. .3862,15765. .15817,19866. .19974,20096. .20167,
                20436. .20605,21110. .21490,21609. .21792,22228. .22448,
                22637. .22752,23118. .23289,23656. .23815,24967. .25111,
                25989. .26120,26318. .26493,26808. .26941,27407. .27525,
                27628. .28683,29126. .29875,30419. .31024,31244. .31393,
                32328. .32471,32599. .32878,32972. .33185,33251. .33410,
                33848. .33952,36954. .37098,37701. .38330,38624. .38780,
                38961. .39213,39338. .39533,41028. .41198,41285. .41428,
                41523. .41733,41824. .41950,43610. .43726,43817. .43866,
                44091. .44275,44353. .44446,47384. .47502,47581. .47783,
                47856. .48050,48501. .48640,49002. .49114,49406. .49547,
                49688. .49813,49997. .50171,50419. .50709,50785. .50919,
                51003. .51308,51399. .51866)
                /gene="PKD1"
                /codon_start=1
                /product="polycystic kidney disease-associated protein"
                /protein_id="AAB59488.1"
                /db_xref="GI:790819"
                /translation="MPPAAPARLALALGLGLWLGALAGGPGRGCGPCEPPCLCGPAPG
                AACRVNCSGRGLRTLGPALRIPADATALTAMAKADGDVKKYGSKRGALALTHGHSLLR
                DVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLFNLSEINLSGNPFECD
                CGLAWLPRWAEEQQVRVVQPEAATCAGPGSLAGQPLLGIPLLDSGCGATLVGPHGPLA
                SGQLAAFHIAAPLPVTATRWDFGDGSAEVDAAGPAASHRYVLPGRYHVTAVLALGAGS
                ALLGTDVQVEAAPAALELVCPSSVQSDESLDLSIQNRGGSGLEAAYSIVALGEEPARA
                VHPLCPSDTEIFPGNGHCYRLVVEKAAWLQAQEQCQAWAGAALAMVDSPAVQRFLVSR
                VTRSLDVWIGFSTVQGVEVGPAPQGEAFSLESCQNWLPGEPHPATAEHCVRLGPTGWC
                NTDLCSAPHSYVCELQPGGPVQDAENLLVGAPSGDLQGPLTPLAQQDGLSAPHEPVEA
                GMGSVELRRPPAHQVMVFPGLRLSREAFLTTAEFGTQELRRPAQLRLQVYRLLSTAGT
                PENGSEPESRSPDNRTQLAPACMPGGRWCPGANICLPLDASCHPQACANGCTEGEHVV
                DVVVENSASRANLSLRVTAEEPICGLRATPSPEARVLQGVLVRYSPVVEAGSDMVFRW
                TINDKQSLTFQNVVFNVIYQSAAVFKLSLTASNHVSNVTVNYNVTVERMNRMQGLQVS
                TVPAVLSPNATLALTAGVLVDSAVEVAFLWTFGDGEQALHQFQPPYNESFPVPDPSVA
                QVLVEHNVMHTYAAPGEYLLTVLASNAFENRTQQVPVSVRASLPSVAVGVSDGVLSQP
                AANHTYASRGTYHVRLEVNNTVSGAAAQADVRVFEELRGLSVDMSLAVEQGAPVVVSA
                AVQTGDNITWTFDMGDGTVLSGPEATVEHVYLRAQNCTVTVGAASPAGHLARSLHVLV
                FVLEVLRVEPAACIPTQPDARLTAYVTGNPAHYLFDWTFGDGSSNTTVRGCPTVTHNF
                TRSGTFPLALVLSSRVNRAHYFTSICVEPEVGNVTLQPERQFVQLGDEAWLVACAWPP
                FPYRYTWDFGTEEAAPTRARGPEVTFIYRDPGSYLVTVTASNNISAANDSALVEVQEP
                VLVTSIKVNGSLGLELQQPYLFSAVGRGRPASYLWDLGDGGWLEGPEVTHAYNSTGDF
                TGPGPGRQRQRLLAHRLEAGTYHVQLRATNMLGSAWADCTMDFVEPVGWLMVAASPNP
                AAVNKSVTLSAELAGGSGVVYTWSLEEGLSWETSEPFTTHSFPTPGLHLVTMTAGNPL
                GSANATVEVDVQVPVSGLSIRASEPGGSFVAAGSSVPFWGQLATGTNVSWCWAVPGGS
                SKRGPHVTMVFPDAGTFSIRLNASNAVSWVSATYNLTAEEPIVGLVLWASSKVVAPGQ
                LVHFQILLAAGSAVTFRLQDTDEPRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTVQ
                VLACREPEVDVVLPLQVLMRRSQRNYLEAHVDLRDCVTYQTEYRWEVYRTASCQRPGR
                PARVALPGVDVSRPRLVLPRLALPVGHYCFVFVVSFGDTPLTQSIQANVTVAPERLVP
                IIEGGSYRVWSDTRDLVLDGSESYDPNLEDGDQTPLSFHWACVASTQREAGCCALNFG
                PRGSSTVTIPRERLAAGVEYTFSLTVWKAGRKEEATNQTVLIRSGRVPIVSLECVSCK
                AQAVYEVSRSSYVYLEGRCLNCSSGSKRGRWAARTFSNKTLVLDETTTSTGSAGMRLV
                LRRGVLRDGEGYTFTLTVLGRSGEEEGCASIRLSPNRPPLGGSCRLFPLGAVHALTTK
                VHFECTGWHDAEDAGAPLVYALLLRRCRQGHCEEFCVYKGSLSSYGAVLPPGFRPHFE
                VGLAVVVQDQLGAAVVALNRSLAITLPEPNGSATGLTVWLHGLTASVLPGLLRQADPQ
                HVIEYSLALVTVLNEHRAQIRKNITETLVSLRVHTVDDIQQIAAALAQCMGPSRELVC
                RSCLKQTLHKLEAMMLILQAETTAGTVTPTAIGDSILNITGDLIHLASSDVRAPQPSE
                LGAESPSRMVASQAYNLTSALMRILMRSRVLNEEPLTLAGEEIVAQGKRSDPRSLLCY
                GGAPGPGCHFSIPEAFSGALANLSDVVQLIFLVDSNPFPFGYISNYTVSTKVASMAFQ
                TQAGAQIPIERLASERAITVKVPNNSDWAARGHRSSANSANSVVVQPQASVGAVVTLD
                SSNPAAGLHLQLNYTLLDGHYLSEEPEPYLAVYLHSEPRPNEHNCSASRRIRPESLQG
                ADHRPYTFFISPGSRDPAGSYHLNLSSHFRWSALQVSVGLYTSLCQYFSEEDMVWRTE
                GLLPLEETSPRQAVCLTRHLTAFGASLFVPPSHVRFVFPEPTADVNYIVMLTCAVCLV
                TYMVMAAILHKLDQLDASRGRAIPFCGQRGRFKYEILVKTGWGRGSGTTAHVGIMLYG
                VDSRSGHRHLDGDRAFHRNSLDIFRIATPHSLGSVWKIRVWHDNKGLSPAWFLQHVIV
                RDLQTARSAFFLVNDWLSVETEANGGLVEKEVLAASDAALLRFRRLLVAELQRGFFDK
                HIWLSIWDRPPRSRFTRIQRATCCVLLICLFLGANAVWYGAVGDSAYSTGHVSRLSPL
                SVDTVAVGLVSSVVVYPVYLAILFLFRMSRSKVAGSPSPTPAGQQVLDIDSCLDSSVL
                DSSFLTFSGLHAEAFVGQMKSDLFLDDSKSLVCWPSGEGTLSWPDLLSDPSIVGSNLR
                QLARGQAGHGLGPEEDGFSLASPYSPAKSFSASDEDLIQQVLAEGVSSPAPTQDTHME
                TDLLSSLSSTPGEKTETLALQRLGELGPPSPGLNWEQPQAARLSRTGLVEGLRKRLLP
                AWCASLAHGLSLLLVAVAVAVSGWVGASFPPGVSVAWLLSSSASFLASFLGWEPLKVL
                LEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVPRVRPPHGFALFLAKEEARKVKRL
                HGMLRSLLVYMLFLLVTLLASYGDASCHGHAYRLQSAIKQELHSRAFLAITRSEELWP
                WMAHVLLPYVHGNQSSPELGPPRLRQVRLQEALYPDPPGPRVHTCSAAGGFSTSDYDV
                GWESPHNGSGTWAYSAPDLLGAWSWGSCAVYDSGGYVQELGLSLEESRDRLRFLQLHN
                WLDNRSRAVFLELTRYSPAVGLHAAVTLRLEFPAAGRALAALSVRPFALRRLSAGLSL
                PLLTSVCLLLFAVHFAVAEARTWHREGRWRVLRLGAWARWLLVALTAATALVRLAQLG
                AADRQWTRFVRGRPRRFTSFDQVAQLSSAARGLAASLLFLLLVKAAQQLRFVRQWSVF
                GKTLCRALPELLGVTLGLVVLGVAYAQLAILLVSSCVDSLWSVAQALLVLCPGTGLST
                LCPAESWHLSPLLCVGLWALRLWGALRLGAVILRWRYHALRGELYRPAWEPQDYEMVE
                LFLRRLRLWMGLSKVKEFRHKVRFEGMEPLPSRSSRGSKVSPDVPPPSAGSDASHPST
                SSSQLDGLSVSLGRLGTRCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQQLHSLQ
                GRRSSRAPAGSSRGPSPGLRPALPSRLARASRGVDLATGPSRTPLRAKNKVHPSST"
     intron     3863. .15764
                /gene="PKD1"
                /number=1
     repeat_region complement(6251. .6540)
                /rpt_family="Alu"
     repeat_region complement(6758. .7024)
                /rpt_family="Alu"
     repeat_region complement(7082. .7371)
                /rpt_family="Alu"
     repeat_region complement(7510. .7763)
                /rpt_family="Alu"
     repeat_region complement(7821. .8107)
                /rpt_family="Alu"
     repeat_region 9490. .9774
                /rpt_family="Alu"
     repeat_region 10178. .10411
                /rpt_family="Alu"
     misc_feature 11958. .12187
                /note="CpG island; putative"
     exon       15765. .15817
                /gene="PKD1"
                /note="putative"
                /number=2
     intron     15818. .19865
                /gene="PKD1"
                /number=2
     repeat_region complement(17907. .18198)
                /rpt_family="Alu"
     repeat_region complement(18260. .18516)
                /rpt_family="Alu"
     repeat_region complement(18574. .18843)
                /rpt_family="Alu"
     repeat_region complement(19064. .19350)
                /rpt_family="Alu"
     exon       19866. .19974
                /gene="PKD1"
                /number=3
     intron     19975. .20095
                /gene="PKD1"
                /number=3
     exon       20096. .20167
                /gene="PKD1"
                /number=4
     intron     20168. .20435
                /gene="PKD1"

Query Match          6.6%;  Score 42;  DB 93;  Length 53522;
Best Local Similarity  52.9%;  Pred. No. 0.27;
Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;
```

```
Qy    469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
          ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db  34691 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34632

Qy    529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
          | | || |  |||| |  ||  | | || |||| | || |         | | ||||
Db  34631 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGAGGGGAGGAGGGGGACAG 34572

Qy    589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          | | ||  ||  ||||| ||||||  ||  | ||  ||| || | |||
Db  34571 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34522


RESULT  8
AR097326/c
LOCUS       AR097326    53526 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 2 from patent US 6071717.
ACCESSION   AR097326
VERSION     AR097326.1  GI:12806056
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 53526)
  AUTHORS   Klinger,K., Burn,T., Connors,T., Dackowski,W., Germino,G., Qian,F.
            and Landes,G.
  TITLE     Polycystic kidney disease gene and protein
  JOURNAL   Patent: US 6071717-A 2 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .53526
                     /organism="unknown"
BASE COUNT     8485 a  17665 c  15769 g  11607 t
ORIGIN


  Query Match             6.6%;  Score 42;  DB 9;  Length 53526;
  Best Local Similarity   52.9%;  Pred. No. 0.27;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
          ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db  34695 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34636

Qy    529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
          | | || |  |||| |  ||  | | || |||| | || |         | | ||||
Db  34635 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGGGGACAG 34576

Qy    589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          | | ||  ||  ||||| ||||||  ||  | ||  ||| || | |||
Db  34575 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34526


RESULT  9
AR097325/c
LOCUS       AR097325    53577 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6071717.
ACCESSION   AR097325
VERSION     AR097325.1  GI:12806055
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 53577)
  AUTHORS   Klinger,K., Burn,T., Connors,T., Dackowski,W., Germino,G., Qian,F.
            and Landes,G.
  TITLE     Polycystic kidney disease gene and protein
  JOURNAL   Patent: US 6071717-A 1 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .53577
                     /organism="unknown"
BASE COUNT     8495 a  17682 c  15784 g  11616 t
ORIGIN


  Query Match             6.6%;  Score 42;  DB 9;  Length 53577;
  Best Local Similarity   52.9%;  Pred. No. 0.27;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
          ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db  34690 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34631

Qy    529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
          | | || |  |||| |  ||  | | || |||| | || |         | | ||||
Db  34630 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGGGGACAG 34571

Qy    589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          | | ||  ||  ||||| ||||||  ||  | ||  ||| || | |||
Db  34570 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34521


RESULT  10
AC015919
LOCUS       AC015919   169713 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone CTD-2047D24 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC015919
VERSION     AC015919.2  GI:9176716
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2047D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6446748.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L855
                Center clone name: 2047_D_24
            --------------.
            * NOTE: This record contains 177 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
```

```
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      841: contig of 841 bp in length
*      842 941: gap of      100 bp
*      942     1778: contig of 837 bp in length
*     1779 1878: gap of      100 bp
*     1879     2729: contig of 851 bp in length
*     2730 2829: gap of      100 bp
*     2830     3681: contig of 852 bp in length
*     3682 3781: gap of      100 bp
*     3782     4620: contig of 839 bp in length
*     4621 4720: gap of      100 bp
*     4721     5544: contig of 824 bp in length
*     5545 5644: gap of      100 bp
*     5645     6505: contig of 861 bp in length
*     6506 6605: gap of      100 bp
*     6606     7451: contig of 846 bp in length
*     7452 7551: gap of      100 bp
*     7552     8385: contig of 834 bp in length
*     8386 8485: gap of      100 bp
*     8486     9315: contig of 830 bp in length
*     9316 9415: gap of      100 bp
*     9416    10262: contig of 847 bp in length
*    10263 10362: gap of      100 bp
*    10363    11228: contig of 866 bp in length
*    11229 11328: gap of      100 bp
*    11329    12180: contig of 852 bp in length
*    12181 12280: gap of      100 bp
*    12281    13121: contig of 841 bp in length
*    13122 13221: gap of      100 bp
*    13222    14070: contig of 849 bp in length
*    14071 14170: gap of      100 bp
*    14171    15030: contig of 860 bp in length
*    15031 15130: gap of      100 bp
*    15131    15974: contig of 844 bp in length
*    15975 16074: gap of      100 bp
*    16075    16900: contig of 826 bp in length
*    16901 17000: gap of      100 bp
*    17001    17835: contig of 835 bp in length
*    17836 17935: gap of      100 bp
*    17936    18775: contig of 840 bp in length
*    18776 18875: gap of      100 bp
*    18876    19721: contig of 846 bp in length
*    19722 19821: gap of      100 bp
*    19822    20666: contig of 845 bp in length
*    20667 20766: gap of      100 bp
*    20767    21625: contig of 859 bp in length
*    21626 21725: gap of      100 bp
*    21726    22617: contig of 892 bp in length
*    22618 22717: gap of      100 bp
*    22718    23551: contig of 834 bp in length
*    23552 23651: gap of      100 bp
*    23652    24492: contig of 841 bp in length
*    24493 24592: gap of      100 bp
*    24593    25450: contig of 858 bp in length
*    25451 25550: gap of      100 bp
*    25551    26411: contig of 861 bp in length
*    26412 26511: gap of      100 bp
*    26512    27357: contig of 846 bp in length
*    27358 27457: gap of      100 bp
*    27458    28297: contig of 840 bp in length
*    28298 28397: gap of      100 bp
*    28398    29223: contig of 826 bp in length
*    29224 29323: gap of      100 bp
*    29324    30171: contig of 848 bp in length
*    30172 30271: gap of      100 bp
*    30272    31100: contig of 829 bp in length
*    31101 31200: gap of      100 bp
*    31201    32044: contig of 844 bp in length
*    32045 32144: gap of      100 bp
*    32145    32989: contig of 845 bp in length
*    32990 33089: gap of      100 bp
*    33090    33951: contig of 862 bp in length
*    33952 34051: gap of      100 bp
*    34052    34908: contig of 857 bp in length
*    34909 35008: gap of      100 bp
*    35009    35851: contig of 843 bp in length
*    35852 35951: gap of      100 bp
*    35952    36798: contig of 847 bp in length
*    36799 36898: gap of      100 bp
*    36899    37756: contig of 858 bp in length
*    37757 37856: gap of      100 bp
*    37857    38702: contig of 846 bp in length
*    38703 38802: gap of      100 bp
*    38803    39635: contig of 833 bp in length
*    39636 39735: gap of      100 bp
*    39736    40577: contig of 842 bp in length
*    40578 40677: gap of      100 bp
*    40678    41508: contig of 831 bp in length
*    41509 41608: gap of      100 bp
*    41609    42452: contig of 844 bp in length
*    42453 42552: gap of      100 bp
*    42553    43405: contig of 853 bp in length
*    43406 43505: gap of      100 bp
*    43506    44360: contig of 855 bp in length
*    44361 44460: gap of      100 bp
*    44461    45294: contig of 834 bp in length
*    45295 45394: gap of      100 bp
*    45395    46256: contig of 862 bp in length
*    46257 46356: gap of      100 bp
*    46357    47221: contig of 865 bp in length
*    47222 47321: gap of      100 bp
*    47322    48156: contig of 835 bp in length
*    48157 48256: gap of      100 bp
*    48257    49090: contig of 834 bp in length
*    49091 49190: gap of      100 bp
*    49191    50043: contig of 853 bp in length
*    50044 50143: gap of      100 bp
*    50144    50989: contig of 846 bp in length
*    50990 51089: gap of      100 bp
*    51090    51925: contig of 836 bp in length
*    51926 52025: gap of      100 bp
*    52026    52874: contig of 849 bp in length
*    52875 52974: gap of      100 bp
*    52975    53824: contig of 850 bp in length
*    53825 53924: gap of      100 bp
*    53925    54787: contig of 863 bp in length
*    54788 54887: gap of      100 bp
*    54888    55740: contig of 853 bp in length
*    55741 55840: gap of      100 bp
*    55841    56692: contig of 852 bp in length
*    56693 56792: gap of      100 bp
*    56793    57654: contig of 862 bp in length
*    57655 57754: gap of      100 bp
*    57755    58595: contig of 841 bp in length
*    58596 58695: gap of      100 bp
*    58696    59533: contig of 838 bp in length
*    59534 59633: gap of      100 bp
*    59634    60479: contig of 846 bp in length
*    60480 60579: gap of      100 bp
*    60580    61430: contig of 851 bp in length
*    61431 61530: gap of      100 bp
*    61531    62371: contig of 841 bp in length
*    62372 62471: gap of      100 bp
*    62472    63313: contig of 842 bp in length
*    63314 63413: gap of      100 bp
*    63414    64260: contig of 847 bp in length
*    64261 64360: gap of      100 bp
*    64361    65225: contig of 865 bp in length
*    65226 65325: gap of      100 bp
*    65326    66181: contig of 856 bp in length
*    66182 66281: gap of      100 bp
*    66282    67127: contig of 846 bp in length
*    67128 67227: gap of      100 bp
```

```
          *    67228    68088: contig of 861 bp in length
          *    68089 68188: gap of      100 bp

  Query Match              6.4%;  Score 41.2;  DB 64;  Length 169713;
  Best Local Similarity   45.9%;  Pred. No. 0.51;
  Matches  133;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy     348 tgtataggaggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaa 407
           ||   ||||  | ||||  |  ||   ||  |  |       |       |  |  | |
Db  165020 TGGGGAGGAAACCAGAGCTAAGACAACTAGTAGATCATAGTAAGAACTTTGGCTTTTTTT 165079

Qy     408 ttactgttcttcatttcaattattcttacctactattcagttcccttgatcttttcttct 467
           || || || ||| ||||  || ||||    | || | |  |||| |   || ||| || |
Db  165080 TTTCTTTTTTTCTTTTCTTTTCTTCTCTTTTCCTTTCCCCTTCCTTCCTTCCTTTTTTTT 165139

Qy     468 tgggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgag 527
           |    |||     |  |   ||| |   |  | |  |      |     || |  |    | |
Db  165140 TTTTTGCTCAAAAAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 165199

Qy     528 acatggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacg 587
           |||| |  || |  |  || ||   ||  ||  |  |          || | ||  |
Db  165200 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 165259

Qy     588 tgtgagggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
           | | ||| ||  ||| |  ||| |||  ||.   |   ||  || ||||
Db  165260 GGGGGGGGGGGGGGGGGGGGGGNGGGNNGGNGGGGGGGGGGGGGGGGGG 165309


RESULT  11
AC004965
LOCUS       AC004965   170758 bp    DNA             PRI       30-SEP-2000
DEFINITION  Homo sapiens PAC clone RP5-1106H14 from 7, complete sequence.
ACCESSION   AC004965
VERSION     AC004965.2  GI:9581948
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170758)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 170758)
  AUTHORS   Bemis,G., Kalicki,J. and Langston,Y.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1106H14
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 170758)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 170758)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 170758)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 170758)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 29, 2000 this sequence version replaced gi:3213032.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_DJ1106H14
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-132A1; the clone sequenced
            to the right is RP5-1059M17, 200 bp overlap.  Actual start of this
            clone is at base position 1 of RP5-1106H14; actual end is at base
            position 58064 of RP5-1059M17.

            RP5-1106H14 contains a single stranded region from 112607 to
            112734.
FEATURES                Location/Qualifiers
     source             1. .170758
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7"
                        /clone="RP5-1106H14"
                        /clone_lib="RPCI-5"
     repeat_region      1. .232
                        /rpt_family="Alu"
     repeat_region      235. .534
                        /rpt_family="Alu"
     repeat_region      1070. .1378
                        /rpt_family="Alu"
     repeat_region      1392. .1527
                        /rpt_family="Alu"
     repeat_region      1654. .1946
                        /rpt_family="Alu"
     repeat_region      2115. .2296
                        /rpt_family="Alu"
     repeat_region      2319. .2495
                        /rpt_family="Alu"
     repeat_region      2643. .2754
                        /rpt_family="Alu"
```

```
repeat_region   2758. .2836
                /rpt_family="Alu"
repeat_region   3701. .3769
                /rpt_family="L2"
repeat_region   3771. .4212
                /rpt_family="ERV1"
repeat_region   4213. .4510
                /rpt_family="Alu"
repeat_region   4514. .4776
                /rpt_family="Alu"
repeat_region   4777. .4897
                /rpt_family="ERV1"
repeat_region   4917. .5017
                /rpt_family="L2"
repeat_region   5253. .5655
                /rpt_family="MaLR"
repeat_region   5656. .6003
                /rpt_family="MaLR"
repeat_region   6009. .6083
                /rpt_family="MaLR"
repeat_region   6261. .6566
                /rpt_family="Alu"
repeat_region   6570. .6984
                /rpt_family="L1"
repeat_region   6985. .7285
                /rpt_family="Alu"
repeat_region   7286. .7541
                /rpt_family="L1"
repeat_region   7573. .7754
                /rpt_family="Alu"
repeat_region   7755. .8054
                /rpt_family="Alu"
repeat_region   8055. .8189
                /rpt_family="Alu"
repeat_region   8219. .8451
                /rpt_family="Alu"
repeat_region   8452. .8747
                /rpt_family="Alu"
repeat_region   8748. .8840
                /rpt_family="Alu"
repeat_region   8891. .9428
                /rpt_family="ERVL"
repeat_region   9607. .9890
                /rpt_family="Alu"
repeat_region   9920. .10230
                /rpt_family="Alu"
repeat_region   10479. .10691
                /rpt_family="Alu"
repeat_region   11763. .11832
                /rpt_family="Alu"
repeat_region   11985. .11999
                /rpt_family="L1"
repeat_region   12000. .12283
                /rpt_family="Alu"
repeat_region   12284. .12411
                /rpt_family="L1"
repeat_region   12412. .12713
                /rpt_family="Alu"
repeat_region   12714. .12786
                /rpt_family="L1"
repeat_region   12787. .13087
                /rpt_family="Alu"
repeat_region   13088. .13099
                /rpt_family="L1"
misc_feature    13089. .13327
                /note="similar to  EST AA307591 (NID:g1959921)"
misc_feature    13089. .13351
                /note="similar to  EST AA243241 (NID:g1874034) zr26e06.r1"
misc_feature    13089. .13309
                /note="similar to  EST AA157629 (NID:g1729237) zo44e07.r1"
misc_feature    13089. .13350
                /note="similar to  EST AI904224 (NID:g6494611)"
misc_feature    13091. .13335
                /note="similar to  EST AA092186 (NID:g1637175)"
misc_feature    13105. .13336
                /note="similar to  EST H49617 (NID:g989458) yq20b08.s1"
repeat_region   13362. .13598
                /rpt_family="L1"
repeat_region   13599. .13766
                /rpt_family="Alu"
repeat_region   13767. .14072
                /rpt_family="Alu"
repeat_region   14073. .14158
                /rpt_family="Alu"

  Query Match             6.4%;  Score 41.2;  DB 85;  Length 170758;
  Best Local Similarity   60.9%;  Pred. No. 0.51;
  Matches   67;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     528 acatggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacg 587
           | | || |||  |||| || ||||||||  | |||| |||||  |     | ||||||
Db   25856 AGAGGGAGAGGAGGAGGTGGAGGAGGAAGAGGAGGAAGAAGATGTTGGAGGCTGAGGAGG 25915

Qy     588 tgtgagggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
             |  || || |  || ||| |    ||  |   | |||  || |||||
Db   25916 AGGAAGAGTAGGAGGTAGAGGAAGAGGAGGAGAAGGAAGAGAAGGAGGAG 25965


RESULT  12
AC015996/c
LOCUS       AC015996   143255 bp    DNA             HTG       20-FEB-2000
DEFINITION  Homo sapiens clone RP11-14A9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC015996
VERSION     AC015996.2  GI:7008798
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143255)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-14A9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143255)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 20, 2000 this sequence version replaced gi:6449544.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3384
                Center clone name: 14_A_9
            --------------
            * NOTE: This record contains 171 individual
            * sequencing reads that have not been assembled into
```

```
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      840: contig of 840 bp in length
*                   gap of unknown length
*      841     1737: contig of 897 bp in length
*                   gap of unknown length
*     1738     2629: contig of 892 bp in length
*                   gap of unknown length
*     2630     3479: contig of 850 bp in length
*                   gap of unknown length
*     3480     4336: contig of 857 bp in length
*                   gap of unknown length
*     4337     5216: contig of 880 bp in length
*                   gap of unknown length
*     5217     6083: contig of 867 bp in length
*                   gap of unknown length
*     6084     6983: contig of 900 bp in length
*                   gap of unknown length
*     6984     7862: contig of 879 bp in length
*                   gap of unknown length
*     7863     8768: contig of 906 bp in length
*                   gap of unknown length
*     8769     9670: contig of 902 bp in length
*                   gap of unknown length
*     9671    10530: contig of 860 bp in length
*                   gap of unknown length
*    10531    11433: contig of 903 bp in length
*                   gap of unknown length
*    11434    12325: contig of 892 bp in length
*                   gap of unknown length
*    12326    13221: contig of 896 bp in length
*                   gap of unknown length
*    13222    14089: contig of 868 bp in length
*                   gap of unknown length
*    14090    14974: contig of 885 bp in length
*                   gap of unknown length
*    14975    15871: contig of 897 bp in length
*                   gap of unknown length
*    15872    16755: contig of 884 bp in length
*                   gap of unknown length
*    16756    17641: contig of 886 bp in length
*                   gap of unknown length
*    17642    18500: contig of 859 bp in length
*                   gap of unknown length
*    18501    19398: contig of 898 bp in length
*                   gap of unknown length
*    19399    20293: contig of 895 bp in length
*                   gap of unknown length
*    20294    21157: contig of 864 bp in length
*                   gap of unknown length
*    21158    22063: contig of 906 bp in length
*                   gap of unknown length
*    22064    22957: contig of 894 bp in length
*                   gap of unknown length
*    22958    23858: contig of 901 bp in length
*                   gap of unknown length
*    23859    24732: contig of 874 bp in length
*                   gap of unknown length
*    24733    25612: contig of 880 bp in length
*                   gap of unknown length
*    25613    26496: contig of 884 bp in length
*                   gap of unknown length
*    26497    27397: contig of 901 bp in length
*                   gap of unknown length
*    27398    28295: contig of 898 bp in length
*                   gap of unknown length
*    28296    29173: contig of 878 bp in length
*                   gap of unknown length
*    29174    30071: contig of 898 bp in length
*                   gap of unknown length
*    30072    30976: contig of 905 bp in length
*                   gap of unknown length
*    30977    31882: contig of 906 bp in length
*                   gap of unknown length
*    31883    32780: contig of 898 bp in length
*                   gap of unknown length
*    32781    33660: contig of 880 bp in length
*                   gap of unknown length
*    33661    34544: contig of 884 bp in length
*                   gap of unknown length
*    34545    35422: contig of 878 bp in length
*                   gap of unknown length
*    35423    36294: contig of 872 bp in length
*                   gap of unknown length
*    36295    37161: contig of 867 bp in length
*                   gap of unknown length
*    37162    38046: contig of 885 bp in length
*                   gap of unknown length
*    38047    38943: contig of 897 bp in length
*                   gap of unknown length
*    38944    39821: contig of 878 bp in length
*                   gap of unknown length
*    39822    40724: contig of 903 bp in length
*                   gap of unknown length
*    40725    41524: contig of 800 bp in length
*                   gap of unknown length
*    41525    42428: contig of 904 bp in length
*                   gap of unknown length
*    42429    43297: contig of 869 bp in length
*                   gap of unknown length
*    43298    44181: contig of 884 bp in length
*                   gap of unknown length
*    44182    45034: contig of 853 bp in length
*                   gap of unknown length
*    45035    45922: contig of 888 bp in length
*                   gap of unknown length
*    45923    46818: contig of 896 bp in length
*                   gap of unknown length
*    46819    47715: contig of 897 bp in length
*                   gap of unknown length
*    47716    48608: contig of 893 bp in length
*                   gap of unknown length
*    48609    49477: contig of 869 bp in length
*                   gap of unknown length
*    49478    50381: contig of 904 bp in length
*                   gap of unknown length
*    50382    51251: contig of 870 bp in length
*                   gap of unknown length
*    51252    52145: contig of 894 bp in length
*                   gap of unknown length
*    52146    53042: contig of 897 bp in length
*                   gap of unknown length
*    53043    53940: contig of 898 bp in length
*                   gap of unknown length
*    53941    54805: contig of 865 bp in length
*                   gap of unknown length
*    54806    55676: contig of 871 bp in length
*                   gap of unknown length
*    55677    56569: contig of 893 bp in length
*                   gap of unknown length
*    56570    57455: contig of 886 bp in length
*                   gap of unknown length
*    57456    58312: contig of 857 bp in length
*                   gap of unknown length
*    58313    59194: contig of 882 bp in length
*                   gap of unknown length
*    59195    60086: contig of 892 bp in length
*                   gap of unknown length
*    60087    60940: contig of 854 bp in length
```

```
            *                    gap of unknown length
            *    60941    61835: contig of 895 bp in length
            *                    gap of unknown length
            *    61836    62731: contig of 896 bp in length
            *                    gap of unknown length
            *    62732    63619: contig of 888 bp in length
            *                    gap of unknown length
            *    63620    64445: contig of 826 bp in length

  Query Match             6.4%;  Score 41;  DB 64;  Length 143255;
  Best Local Similarity   52.4%;  Pred. No. 0.58;
  Matches   89;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagcccctgaga 528
           | |||  |    ||||   |||||   |  |||| |   |    ||| |       | |
Db   87123 GAGGGGAGGGAGAGGGAAGGGGAGGGGGGGGAAGAAGGGGGGGGAGGGGGGAGGGGGGGG 87064

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
           ||||  |  || |  |  |||||  |   ||| | |||        || | ||| |
Db   87063 AGGGGGGGGGGGGGGGGGGAGAGGAGGGAGGGGAGGGAGAGGGGGGGGGGGGGAGGA 87004

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
           | | ||  ||  ||||| ||||||| | ||    | | ||| || |||||
Db   87003 GGGGGGAGGGGGGGGAGGAGGGAGGNGGGGGGGGGGAGGGAAGGGGGGGA 86954


RESULT 13
AC006013/c
LOCUS       AC006013   159002 bp    DNA             PRI       27-APR-2000
DEFINITION  Homo sapiens PAC clone RP5-945F2 from 7p11.2-q11.2, complete
            sequence.
ACCESSION   AC006013
VERSION     AC006013.3  GI:6358869
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159002)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 159002)
  AUTHORS   Scott,K., Wohldmann,P. and Krauss,B.
  TITLE     The sequence of Homo sapiens PAC clone RP5-945F2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 159002)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159002)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159002)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 11, 1999 this sequence version replaced gi:4337281.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_DJ0945F02
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP5-945F2;
            actual end is at base position 159002 of RP5-945F2.
FEATURES             Location/Qualifiers
     source          1. .159002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p11.2-q11.2"
                     /clone="RP5-945F2"
                     /clone_lib="RPCI-5"
     repeat_region   814. .1225
                     /rpt_family="Retroviral"
     repeat_region   1226. .1430
                     /rpt_family="Retroviral"
     repeat_region   1444. .1478
                     /rpt_family="Retroviral"
     repeat_region   1982. .2071
                     /rpt_family="MaLR"
     repeat_region   2245. .2533
                     /rpt_family="Alu"
     repeat_region   2610. .2899
                     /rpt_family="Alu"
     repeat_region   3435. .3740
                     /rpt_family="Alu"
     repeat_region   4008. .4070
                     /rpt_family="Alu"
     repeat_region   4073. .4128
                     /rpt_family="Alu"
     repeat_region   4142. .4372
                     /rpt_family="Alu"
     repeat_region   4679. .4801
                     /rpt_family="Alu"
     misc_feature    4686. .4704
                     /note="similar to EST AI267554 (NID:g3886721) aq90g12.x1"
     misc_feature    4687. .4703
                     /note="similar to EST AI690708 (NID:g4902010) tx88d03.x1"
     misc_feature    4687. .4704
                     /note="similar to EST AI433071 (NID:g4286607) th41a04.x1"
```

```
repeat_region   4915. .5212
                /rpt_family="Alu"
repeat_region   5290. .5335
                /rpt_family="L1"
repeat_region   5336. .5715
                /rpt_family="MaLR"
misc_feature    5713. .5736
                /note="similar to EST AA730936 (NID:g2752140) nw55a04.s1"
repeat_region   5716. .6015
                /rpt_family="Alu"
misc_feature    5719. .5730
                /note="similar to EST AI433071 (NID:g4286607) th41a04.x1"
repeat_region   6016. .6299
                /rpt_family="Alu"
repeat_region   6300. .6360
                /rpt_family="MaLR"
repeat_region   6361. .6470
                /rpt_family="L1"
repeat_region   6471. .6778
                /rpt_family="Alu"
repeat_region   6779. .7194
                /rpt_family="L1"
repeat_region   7195. .7486
                /rpt_family="Alu"
repeat_region   7487. .7590
                /rpt_family="L1"
repeat_region   7757. .8110
                /rpt_family="Retroviral"
repeat_region   8248. .8595
                /rpt_family="L2"
repeat_region   8597. .8707
                /rpt_family="Alu"
repeat_region   8746. .8860
                /rpt_family="Alu"
repeat_region   8897. .8999
                /rpt_family="(TCTA)n"
repeat_region   9000. .9179
                /rpt_family="(TCCA)n"
repeat_region   10024. .10473
                /rpt_family="MER4-group"
repeat_region   11101. .11152
                /rpt_family="MIR"
repeat_region   11156. .11243
                /rpt_family="Retroviral"
repeat_region   11244. .11538
                /rpt_family="Alu"
repeat_region   11539. .11725
                /rpt_family="Retroviral"
misc_feature    11726. .11747
                /note="similar to EST AI569548 (NID:g4532922) to28e06.x1"
repeat_region   11726. .12027
                /rpt_family="Alu"
repeat_region   12028. .12117
                /rpt_family="Retroviral"
repeat_region   12118. .13108
                /rpt_family="MER4-group"
repeat_region   13100. .13188
                /rpt_family="Retroviral"
repeat_region   14193. .14300
                /rpt_family="L1"
repeat_region   14304. .14587
                /rpt_family="Alu"
repeat_region   14685. .14995
                /rpt_family="Alu"
repeat_region   15525. .15545
                /rpt_family="AT_rich"
repeat_region   15546. .15569
                /rpt_family="(GA)n"
repeat_region   16356. .16377
                /rpt_family="(TTTTA)n"
repeat_region   16380. .16654
                /rpt_family="Alu"
repeat_region   16692. .16788
                /rpt_family="MIR"
repeat_region   17205. .17307
                /rpt_family="L2"
repeat_region   17324. .17382
                /rpt_family="MER1_type"
repeat_region   17456. .17591
                /rpt_family="MER1_type"
repeat_region   17660. .17747
                /rpt_family="MIR"
```

```
  Query Match              6.4%;  Score 40.8;  DB 86;  Length 159002;
  Best Local Similarity   53.8%;  Pred. No. 0.67;
  Matches   84;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     483 ggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttgga 542
           ||  | |||||     |||| |  | ||  ||||| ||     |||   | ||||  |||
Db   83079 GGAGAAGGAGAAGAAGGAAGGAGAAAAAGAAGGAGGAGGAAGAAGATGGGAGGAGGAGGA 83020

Qy     543 gctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtggctgg 602
           |  | |||||||  |    |  |||| |      ||| | ||| |   |||||  ||  ||
Db   83019 GAAGGAGGAGGAGGGAGGAGGGGAAGGAGGAGGGAGGAGGGGAAGGAGGAGGGAGGAGGG 82960

Qy     603 gagaagggaggggtggtcacgaatggacggagggga 638
           ||    |||||    |  | ||   |  |||  ||
Db   82959 GAAGGAGGAGGATGAGGGAGGAGGAGGATGAGAAGA 82924
```

```
RESULT  14
AL133273
LOCUS       AL133273    42375 bp    DNA             PRI       20-APR-2000
DEFINITION  Human DNA sequence from clone RP11-518F7 on chromosome Xq26.3-27.3.
            Contains GSSs, complete sequence.
ACCESSION   AL133273
VERSION     AL133273.5  GI:7228186
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 42375)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 10, 2000 this sequence version replaced gi:6983214.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-518F7 is from the library RPCI-11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-518F7 It may be shorter because we sequence overlapping
```

```
          sections only once, except for a 100 base overlap.
          The true left end of clone GS1-82I1 is at 42276 in this sequence.
          The true right end of clone GS1-164F24 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .42375
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q26.3-27.3"
                     /clone="RP11-518F7"
                     /clone_lib="RPCI-11.2"
     repeat_region   375. .1234
                     /note="L1ME3A repeat: matches 5290. .6147 of consensus"
     repeat_region   1368. .1973
                     /note="L1ME1 repeat: matches 5552. .6166 of consensus"
     repeat_region   1973. .2165
                     /note="L1 repeat: matches 3616. .3817 of consensus"
     repeat_region   2162. .2395
                     /note="L1ME1 repeat: matches 5321. .5554 of consensus"
     repeat_region   2403. .3174
                     /note="L1 repeat: matches 2535. .3301 of consensus"
     repeat_region   3232. .4332
                     /note="L1MA9 repeat: matches 5159. .6305 of consensus"
     repeat_region   4368. .5282
                     /note="L1M4 repeat: matches 2407. .3347 of consensus"
     repeat_region   6671. .6710
                     /note="20 copies 2 mer gt 87% conserved"
     repeat_region   6712. .7080
                     /note="MLT1A1 repeat: matches 4. .365 of consensus"
     repeat_region   7274. .7440
                     /note="MIR repeat: matches 15. .185 of consensus"
     repeat_region   7439. .7472
                     /note="L2 repeat: matches 2665. .2738 of consensus"
     repeat_region   8167. .8464
                     /note="L2 repeat: matches 2470. .2744 of consensus"
     repeat_region   8483. .8783
                     /note="AluY repeat: matches 1. .300 of consensus"
     repeat_region   8958. .8981
                     /note="12 copies 2 mer tg 95% conserved"
     repeat_region   9215. .9400
                     /note="MIR repeat: matches 70. .262 of consensus"
     repeat_region   9786. .11597
                     /note="L1MA7 repeat: matches 4345. .6160 of consensus"
     repeat_region   11605. .13837
                     /note="L1PA13 repeat: matches 3899. .6156 of consensus"
     repeat_region   13829. .14542
                     /note="L1PA13 repeat: matches -651. .88 of consensus"
     repeat_region   14670. .15237
                     /note="L1 repeat: matches 3775. .4349 of consensus"
     repeat_region   15256. .15995
                     /note="L1 repeat: matches 2447. .3188 of consensus"
     repeat_region   16031. .16108
                     /note="39 copies 2 mer ta 65% conserved"
     repeat_region   16115. .16348
                     /note="L1 repeat: matches 2221. .2457 of consensus"
     repeat_region   16692. .16747
                     /note="28 copies 2 mer tg 69% conserved"
     repeat_region   16883. .17220
                     /note="Tigger3b repeat: matches 2. .336 of consensus"
     repeat_region   17221. .17531
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   17532. .18337
                     /note="Tigger3b repeat: matches 336. .1134 of consensus"
     repeat_region   18338. .18645
                     /note="AluY repeat: matches 1. .306 of consensus"
     repeat_region   18646. .18719
                     /note="Tigger3b repeat: matches 1134. .1231 of consensus"
     repeat_region   19033. .19137
                     /note="MER5A repeat: matches 87. .189 of consensus"
     repeat_region   19132. .19196
                     /note="L2 repeat: matches 2661. .2732 of consensus"
     repeat_region   19265. .19656
                     /note="L2 repeat: matches 2077. .2509 of consensus"
     repeat_region   19873. .20221
                     /note="L1PA7 repeat: matches 5787. .6143 of consensus"
     repeat_region   20470. .20585
                     /note="MIR repeat: matches 48. .163 of consensus"
     repeat_region   22963. .23018
                     /note="L2 repeat: matches 2648. .2703 of consensus"
     repeat_region   23153. .23282
                     /note="MIR repeat: matches 54. .190 of consensus"
     repeat_region   23853. .24848
                     /note="L1PA11 repeat: matches 4212. .5201 of consensus"
     repeat_region   24822. .25001
                     /note="L1PA11 repeat: matches 5136. .5311 of consensus"
     repeat_region   25002. .25045
                     /note="MER44C repeat: matches 92. .139 of consensus"
     repeat_region   25046. .25825
                     /note="L1PA11 repeat: matches 5311. .6153 of consensus"
     repeat_region   26594. .26705
                     /note="MIR repeat: matches 87. .192 of consensus"
     repeat_region   27084. .27381
                     /note="AluSp repeat: matches 5. .303 of consensus"
     repeat_region   27393. .27538
                     /note="MER5A repeat: matches 1. .168 of consensus"
     repeat_region   27616. .27866
                     /note="L2 repeat: matches 1292. .1586 of consensus"
     repeat_region   27867. .28194
                     /note="MER33 repeat: matches 1. .324 of consensus"
     repeat_region   28195. .28253
                     /note="L2 repeat: matches 1241. .1292 of consensus"
     repeat_region   28534. .28837
                     /note="AluSq repeat: matches 1. .304 of consensus"
     repeat_region   29130. .29263
                     /note="L1 repeat: matches 3095. .3234 of consensus"
     repeat_region   29300. .29600
                     /note="AluSg1 repeat: matches 7. .291 of consensus"
     repeat_region   29959. .30006
                     /note="24 copies 2 mer ta 100% conserved"
     misc_feature    29959. .30001
                     /note="possible slipping in TA repeat"
     repeat_region   30008. .30296
                     /note="AluJb repeat: matches 1. .287 of consensus"
     repeat_region   30435. .32004
                     /note="TIGGER1 repeat: matches 1. .1561 of consensus"
     repeat_region   32005. .32305
                     /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region   32306. .33037
                     /note="TIGGER1 repeat: matches 1561. .2319 of consensus"
     repeat_region   33038. .33398
                     /note="THE1B repeat: matches 1. .362 of consensus"
     repeat_region   33399. .33495
                     /note="TIGGER1 repeat: matches 2319. .2414 of consensus"
     repeat_region   34358. .35160
                     /note="LTR1 repeat: matches 1. .785 of consensus"
     repeat_region   36146. .36249
                     /note="Tigger3(Golem) repeat: matches 1. .104 of consensus"
     repeat_region   36245. .36951
                     /note="Tigger3(Golem) repeat: matches 2277. .3028 of
                     consensus"
     repeat_region   37711. .37935
                     /note="AluSc repeat: matches 83. .308 of consensus"
     repeat_region   38401. .38711
                     /note="L2 repeat: matches 2387. .2702 of consensus"
     misc_feature    complement(39393. .40114)
                     /note="match: GSS: Em:AQ046215"
     misc_feature    complement(39478. .40104)
                     /note="match: GSS: Em:AQ835014"
     repeat_region   39753. .40001
                     /note="AluSq repeat: matches 49. .298 of consensus"
     misc_feature    complement(39824. .40494)
                     /note="match: GSS: Em:AQ474328"
     repeat_region   40350. .40495
                     /note="L2 repeat: matches 2325. .2470 of consensus"
     repeat_region   41233. .41287
                     /note="L2 repeat: matches 2692. .2750 of consensus"
```

```
     repeat_region   41841. .41900
                     /note="30 copies 2 mer tg 81% conserved"
     repeat_region   42045. .42227
                     /note="L2 repeat: matches 2121. .2308 of consensus"
     repeat_region   42234. .42314
                     /note="MER5A repeat: matches 108. .189 of consensus"
BASE COUNT    12892 a   7569 c   8072 g  13842 t
ORIGIN


  Query Match             6.3%;  Score 40.4;  DB 89;  Length 42375;
  Best Local Similarity   58.2%;  Pred. No. 0.79;
  Matches   71;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy     348 tgtataggaggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaa 407
           | | | | |   |  || |||  |   || |     |   | |||  ||||||||| |||
Db   29147 TATTTTGAAATAACCAGTTCTTGGATTTTGTACTTTCTGCTGCTCATCTCTGTATCCCAA 29206

Qy     408 ttactgttcttcatttcaattattcttacctactattcagttcccttgatcttttcttct 467
           || |||||||| | ||  || ||| ||  || ||      | || ||||| || | ||
Db   29207 TTGCTGTTCTTTACTTTTATCATTATTGGCTCCTTCCTCCTGCCTTTGATTTTGTTTTTG 29266

Qy     468 tg 469
           ||
Db   29267 TG 29268


RESULT  15
AC002352/c
LOCUS       AC002352   162405 bp    DNA             PRI       10-NOV-1998
DEFINITION  Homo sapiens 12q24 PAC P256D10 complete sequence.
ACCESSION   AC002352
VERSION     AC002352.1  GI:3858888
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162405)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chang,J.,
            Chen,E., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
            Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
            Kampal,R., Karpathy,S., Kucherlapati,R., Lau,S., Leal,B., Liu,W.,
            Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
            Rashid,N.D., Renault,B., Rowland,K., Savage,L., Scherer,S.S.,
            Shen,H., Timms,K.M., Todd,J., Vo,Q., Worley,K.C., Yu,W.,
            Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162405)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 162405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 162405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 162405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 162405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   7  (bases 1 to 162405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 10, 1998 this sequence version replaced gi:2961442.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .162405
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="P256D10"
                     /chromosome="12"
                     /map="12q24"
     repeat_region   complement(1. .1741)
                     /partial
                     /rpt_family="HERVL"
     misc_feature    complement(1. .206)
                     /note="overlap bases 1. .206 in AC002978"
     repeat_region   complement(1757. .2111)
                     /rpt_family="THE1A"
     repeat_region   complement(2113. .3263)
                     /rpt_family="THE1A-internal"
     repeat_region   complement(3266. .3568)
                     /rpt_family="AluSx"
     repeat_region   complement(3569. .3998)
                     /rpt_family="THE1A-internal"
     repeat_region   complement(3999. .4349)
                     /rpt_family="THE1A"
     repeat_region   complement(4351. .4963)
                     /rpt_family="HERVL"
     repeat_region   complement(5064. .5251)
                     /rpt_family="HERVL"
     repeat_region   complement(5253. .5290)
                     /rpt_family="MLT2CB"
     repeat_region   complement(5291. .5322)
                     /rpt_family="(GA)n"
     repeat_region   complement(5322. .5350)
                     /rpt_family="(CA)n"
     repeat_region   complement(5416. .5571)
                     /rpt_family="L2"
     repeat_region   complement(5752. .6033)
                     /rpt_family="L2"
     repeat_region   6084. .6644
                     /rpt_family="L1MB7"
     repeat_region   complement(6725. .7111)
                     /rpt_family="L2"
     repeat_region   complement(7263. .7474)
                     /rpt_family="MIR"
     repeat_region   complement(7633. .7931)
                     /rpt_family="AluJb"
     repeat_region   complement(8045. .8097)
                     /rpt_family="(CAA)n"
     repeat_region   complement(8209. .8519)
                     /rpt_family="AluJb"
     repeat_region   8656. .8697
                     /rpt_family="MIR"
     repeat_region   9180. .9232
```

```
                     /rpt_family="MIR"
repeat_region        9457. .9745
                     /rpt_family="LTR16A"
repeat_region        9787. .9897
                     /rpt_family="MIR"
repeat_region        10589. .10695
                     /rpt_family="MIR"
repeat_region        11469. .11533
                     /rpt_family="MIR"
repeat_region        11929. .11957
                     /rpt_family="(CA)n"
repeat_region        12037. .12121
                     /rpt_family="MIR"
repeat_region        complement(12143. .12348)
                     /rpt_family="MIR"
repeat_region        12397. .12466
                     /rpt_family="L2"
repeat_region        12506. .12799
                     /rpt_family="AluSx"
repeat_region        14656. .14724
                     /rpt_family="L2"
repeat_region        15164. .15206
                     /rpt_family="L2"
repeat_region        complement(16063. .16367)
                     /rpt_family="AluJo"
repeat_region        complement(16370. .16545)
                     /rpt_family="AluSg"
repeat_region        complement(16548. .16575)
                     /rpt_family="(CAAAA)n"
repeat_region        complement(16576. .16705)
                     /rpt_family="AluSg"
repeat_region        complement(16718. .17145)
                     /rpt_family="L1MC1"
repeat_region        17146. .17173
                     /rpt_family="THE1C"
repeat_region        17167. .17207
                     /rpt_family="MLT1B"
repeat_region        complement(17277. .17342)
                     /rpt_family="MIR"
repeat_region        17427. .17546
                     /rpt_family="MIR"
repeat_region        17553. .17849
                     /rpt_family="AluSx"
repeat_region        18660. .19251
                     /rpt_family="L1MB2"
repeat_region        19415. .19717
                     /rpt_family="AluJo"
repeat_region        complement(19758. .19869)
                     /rpt_family="MIR"
repeat_region        complement(19898. .20055)
                     /rpt_family="MER30"
repeat_region        complement(20084. .20303)
                     /rpt_family="MER20"
repeat_region        20814. .20880
                     /rpt_family="L2"
repeat_region        22705. .23005
                     /rpt_family="AluSx"
repeat_region        complement(23924. .24022)
                     /rpt_family="L2"
repeat_region        24027. .24361
                     /rpt_family="MER7A"
repeat_region        complement(24365. .24423)
                     /rpt_family="L2"
repeat_region        complement(24480. .24857)
                     /rpt_family="MLT1B"
misc_feature         25192. .25570
                     /note="Region: Similar to AA179159 zp12c04.s1 Stratagene
                     fetal retina 937202 Homo sapiens cDNA clone 609222 3'"
repeat_region        25640. .25790
                     /rpt_family="MIR"
repeat_region        25828. .26074
                     /rpt_family="MIR"
repeat_region        26618. .26648
                     /rpt_family="POLY_A"
misc_feature         complement(27034. .27382)
                     /note="Region: Similar to AA179351 zp12c04.r1 Stratagene
                     fetal retina 937202 Homo sapiens cDNA clone 609222 5'"
repeat_region        27398. .27706
                     /rpt_family="AluSq"
repeat_region        complement(27973. .28041)
                     /rpt_family="L2"
repeat_region        complement(28075. .28286)
                     /rpt_family="L2"

  Query Match             6.3%;  Score 40.2;  DB 85;  Length 162405;
  Best Local Similarity   53.5%;  Pred. No. 1;
  Matches   84;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     481 agggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttg 540
           |||   | | |||  |  |  |    |||| | ||||  |     |||  ||||  |  |
Db 128219 AGGAGGAAGAAGAGGGAGGGGGAGGAAGAAGTGGGAGGGGGAGGAAGAAGTGGGAGGGGG 128160

Qy     541 gagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtggct 600
           |  ||||  |||  || |||| ||||          |  || || ||| ||||| |
Db 128159 AGGAAGAAGTGGGAGGGGGAGGAAGAAGTGGGAGGGGGAGGAAGAAGTGGGAGGGGGAGG 128100

Qy     601 gggagaagggaggggtggtcacgaatggacggagggg 637
           ||   |||||||| |  |  | |||  ||||| |
Db 128099 AAGAAGTGGGAGGGGGAGGAAGAAGTGGGAGGAGGAG 128063


Search completed: April 30, 2001, 19:14:51
Job time: 37365 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:06:41 ; Search time 933.78 Seconds
(without alignments)
399.491 Million cell updates/sec

Title: US-09-721-183-4
Perfect score: 639
Sequence: 1 ccagaaccgagtttaggtcc.....cacgaatggacggaggggat 639

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	42	6.6	53526	19	T94101	Human PKD1 gene.
c	2	42	6.6	53577	17	T18551	Human polycystic k
c	3	42	6.6	53577	19	T94108	Human PKD1 locus b
	4	37.4	5.9	1000	21	A02484	Human colon cancer
c	5	37.2	5.8	1038	14	Q51019	EBV VCA-p-40 gene
c	6	36.6	5.7	10732	21	A10594	Gene encoding a su
	7	34.4	5.4	3614	19	V06567	Arabidopsis cellul
c	8	34.2	5.4	1662	21	C76992	Human ORFX ORF2547
c	9	33.8	5.3	4544	18	V74828	Staphylococcus aur
c	10	33.6	5.3	337	18	T62741	Tobacco acetolacta
	11	33.6	5.3	795	19	V55830	FLGA insert stabil
c	12	33.4	5.2	269	20	V89887	EST clone CT320.
c	13	33.4	5.2	422	21	C98193	Human colon cancer
c	14	33.4	5.2	745	21	A87692	Human secreted pro
c	15	33.4	5.2	9312	21	Z43768	Human CRP-ductin c
	16	33.2	5.2	633	21	C47613	Arabidopsis thalia
	17	33.2	5.2	686	21	C40708	Arabidopsis thalia
	18	33.2	5.2	700	21	C43651	Arabidopsis thalia
c	19	33.2	5.2	996	21	C99916	Human secreted pro
c	20	33	5.2	867	21	Z98056	Human secreted pro
	21	32.8	5.1	975	21	C52115	Arabidopsis thalia
	22	32.8	5.1	1138	21	C79743	Human secreted pro
c	23	32.8	5.1	6138	21	Z57864	Protein regulating
	24	32.6	5.1	909	20	X57382	Rat U3 gene trap d
c	25	32.4	5.1	3680	17	T10665	Mouse bradykinin B
	26	32.2	5.0	741	21	A44712	Human secreted exp
c	27	32.2	5.0	2030	21	C63913	Cucumber xylem sap
	28	32.2	5.0	4289	19	V62147	HSV-2 strain SB5 C
c	29	32.2	5.0	29433	21	F20950	Human interleukin-
c	30	32.2	5.0	29433	21	A34828	Human adenosine re
c	31	32.2	5.0	209273	21	F21437	Human factor-relat
c	32	32	5.0	611	18	T83935	DNA encoding a Sta
c	33	32	5.0	1262	18	V74767	Staphylococcus aur
c	34	32	5.0	2577	21	A88560	Human CC chemokine
c	35	32	5.0	2577	22	D02134	Degenerate human a
c	36	32	5.0	2930	12	Q11494	Tobacco SURB-Hra m
c	37	32	5.0	2930	13	Q28387	Gene from the ALS
c	38	32	5.0	2930	16	Q81181	ALS SURB-Hra (P197
c	39	32	5.0	2930	18	T72863	Tobacco S4 gene en
c	40	32	5.0	2946	9	N81458	Hra mutant gene en
c	41	32	5.0	2946	17	T33352	Tobacco herbicide-
	42	31.8	5.0	1264	21	C37723	Arabidopsis thalia
	43	31.8	5.0	1265	21	C50678	Arabidopsis thalia
	44	31.8	5.0	1376	21	C42540	Arabidopsis thalia
	45	31.8	5.0	1378	21	C50677	Arabidopsis thalia

ALIGNMENTS

RESULT 1
T94101/c
ID T94101 standard; DNA; 53526 BP.
XX
AC T94101;
XX
DT 01-JUN-1998 (first entry)
XX
DE Human PKD1 gene.
XX
KW Human; polycystic kidney disease 1; PKD1; treatment;
KW autosomal dominant polycystic kidney disease; APKD; ss.
XX
OS Homo sapiens.
XX
PN WO9744457-A1.
XX
PD 27-NOV-1997.
XX
PF 22-MAY-1997; 97WO-US08799.
XX
PR 03-JUN-1996; 96US-0658136.
PR 24-MAY-1996; 96US-0655360.
XX
PA (GENZ) GENZYME CORP.
XX
PI Burn T, Connors T, Dackowski W, Germino G, Klinger K;
PI Qian F;
XX
DR WPI; 1998-018511/02.
XX
PT Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
XX

```
PS   Claim 2; Pages 90-118; 257pp; English.
XX
CC   The present sequence is the human polycystic kidney disease 1
CC   (PKD1) gene. The PKD1 gene or polypeptide may be used to treat
CC   autosomal dominant polycystic kidney disease (APKD), and identify
CC   carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
CC   Antibodies (Ab) that distinguish between normal and mutant PKD1
CC   sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
CC   also be used to perform subcellular and histochemical localisation
CC   studies, and to block the function of PKD1. Ab are also useful in
CC   rational drug design studies to identify and test inhibitors of
CC   PKD1. Sense and antisense sequences derived from the PKD1 gene may
CC   used for detection and therapy.
XX
SQ   Sequence 53526 BP; 8486 A; 17665 C; 15768 G; 11607 T; 0 other;


  Query Match              6.6%;  Score 42;  DB 19;  Length 53526;
  Best Local Similarity   52.9%;  Pred. No. 0.03;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
          ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db  34695 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34636

Qy    529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
          | | || |  ||||  |  ||  | | || |||| | || |        | | ||||
Db  34635 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGGAGGAGGGGGGACAG 34576

Qy    589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          | | ||  ||   ||||| ||||||   ||  | ||  ||| || | |||
Db  34575 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34526


RESULT   2
T18551/c
ID   T18551 standard; DNA; 53577 BP.
XX
AC   T18551;
XX
DT   06-MAY-1997  (first entry)
XX
DE   Human polycystic kidney disease normal PKD1 gene.
XX
KW   Adult onset polycystic kidney disease; APKD; autosomal dominant;
KW   mutant; transversion; transition; deletion; insertion; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    4379..5272
FT                   /*tag=  a
FT                   /note= "specifically claimed region of intronless
FT                          cDNA identified by exon trapping"
FT   old_sequence    replace(50652..50653, cg)
FT                   /*tag=  b
FT                   /note= "changes Val codon to Leu codon"
FT   old_sequence    replace(50796..50797, cg)
FT                   /*tag=  c
FT                   /note= "replaces Val codon by Leu codon"
FT   old_sequence    insert(51827..51828, cc)
FT                   /*tag=  d
FT                   /note= "insertion, results in frameshift"
XX
PN   WO9612033-A1.
XX
PD   25-APR-1996.
XX
PF   11-OCT-1995;   95WO-US13357.
XX
PR   31-JAN-1995;   95US-0381520.
PR   12-OCT-1994;   94US-0323443.
XX
PA   (IGIG-) IG LAB INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Burn TC,  Connors TD,  Dackowski W,  Germino G,  Klinger KW;
PI   Landes GM,  Qian F;
XX
DR   WPI; 1996-222017/22.
XX
PT   Isolated human polycystic kidney disease gene and its mutants -
PT   useful for treatment of polycystic kidney disease and screening for
PT   carriers
XX
PS   Claim 1; Fig 1; 65pp; English.
XX
CC   The present sequence is that of the normal human PKD1 gene from
CC   chromosome 16. Mutations in this gene (e.g. transitions,
CC   transversions, deletions and/or insertions) are associated with
CC   adult-onset polycystic kidney disease (APKD). The PKD1 locus is
CC   GC-rich (62.4%). Comparison of this sequence with a previously
CC   reported partial cDNA sequence revealed differences at three
CC   locations (see features table). The most significant difference is
CC   the presence of two additional cytosine residues on the plus-strand
CC   at position 4566 of the previously reported sequence. The insertion
CC   results in a frame-shift in the predicted protein coding sequence,
CC   leading to replacement of 92 C-terminal amino acids with a novel
CC   12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
CC   There is a region consisting of 17 tandem copies of a perfect 27 bp
CC   repeat and two large CT-rich regions.
XX
SQ   Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T; 0 other;


  Query Match              6.6%;  Score 42;  DB 17;  Length 53577;
  Best Local Similarity   52.9%;  Pred. No. 0.03;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
          ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db  34691 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34632

Qy    529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
          | | || |  ||||  |  ||  | | || |||| | || |        | | ||||
Db  34631 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGGAGGAGGGGGGACAG 34572

Qy    589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          | | ||  ||   ||||| ||||||   ||  | ||  ||| || | |||
Db  34571 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34522


RESULT   3
T94108/c
ID   T94108 standard; DNA; 53577 BP.
XX
AC   T94108;
XX
DT   01-JUN-1998  (first entry)
XX
DE   Human PKD1 locus between chromosomal markers ATPL (ATP6C) and D16S84.
XX
KW   Human; polycystic kidney disease 1; PKD1; treatment;
KW   autosomal dominant polycystic kidney disease; APKD; ss.
XX
OS   Homo sapiens.
XX
PN   WO9744457-A1.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08799.
XX
PR   03-JUN-1996;   96US-0658136.
PR   24-MAY-1996;   96US-0655360.
XX
```

PA (GENZ) GENZYME CORP.
XX
PI Burn T, Connors T, Dackowski W, Germino G, Klinger K;
PI Qian F;
XX
DR WPI; 1998-018511/02.
XX
PT Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
XX
PS Example 5; Pages 60-89; 257pp; English.
XX
CC The present sequence is the human polycystic kidney disease 1
CC (PKD1) locus between chromosomal markers ATPL (ATP6C) and D16S84.
CC The PKD1 gene or polypeptide may be used to treat autosomal
CC dominant polycystic kidney disease (APKD), and identify carriers
CC of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC (Ab) that distinguish between normal and mutant PKD1 sequences can
CC also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC perform subcellular and histochemical localisation studies, and to
CC block the function of PKD1. Ab are also useful in rational drug
CC design studies to identify and test inhibitors of PKD1. Sense and
CC antisense sequences derived from the PKD1 gene may used for
CC detection and therapy.
XX
SQ Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T; 0 other;

Query Match 6.6%; Score 42; DB 19; Length 53577;
Best Local Similarity 52.9%; Pred. No. 0.03;
Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
Db 34690 GCGCAGGGGAGGAGGGGAGGACGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34631

Qy 529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
Db 34630 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGGGGGACAG 34571

Qy 589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
Db 34570 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34521

RESULT 4
A02484
ID A02484 standard; cDNA; 1000 BP.
XX
AC A02484;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2475.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPI; 2000-126369/11.
XX
PT Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1; Page 994; 1097pp; English.
XX
CC A00010 to A02716 represent polynucleotides isolated from cDNA libraries
CC constructed from human colon cancer cell lines. The present invention
CC also describes a method of detecting differentially expressed genes
CC correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 1000 BP; 21 A; 53 C; 574 G; 22 T; 330 other;

Query Match 5.9%; Score 37.4; DB 21; Length 1000;
Best Local Similarity 42.0%; Pred. No. 0.11;
Matches 71; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
Db 299 gggggnggggnggggggggggnggggggcagnggggnggcgngngnngngnnnggnang 358

Qy 529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
Db 359 gggggggnggggggngnggnggggnnngnngggggggggaggnnnngnggnggngggggg 418

Qy 589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
Db 419 gnggggggngcnnggnggggggngggggggnnggggngnggaggcgnggg 467

RESULT 5
Q51019/c
ID Q51019 standard; DNA; 1038 BP.
XX
AC Q51019;
XX
DT 22-MAY-1994 (first entry)
XX
DE EBV VCA-p-40 gene ORF BdRF1.
XX
KW Viral capsid antigen; Epstein Barr Virus; viral genome; antibodies;
KW diagnosis; ss.
XX
OS Epstein-barr virus.
XX
PN AU9335152-A.
XX
PD 16-SEP-1993.
XX

```
PF   12-MAR-1993;   93AU-0035152.
XX
PR   13-MAR-1992;   92EP-0200721.
XX
PA   (ALKU ) AKZO NV.
XX
PI   Middeldorp JM;
XX
DR   WPI; 1993-345368/44.
DR   P-PSDB; R42317.
XX
PT   Peptide(s) and nucleic acid sequences related to EBV - for
PT   detecting EBV in samples
XX
PS   Claim 8; Page 39; 62pp; English.
XX
CC   The sequence is that of Epstein Barr Virus open reading frame BdRF1
CC   which encodes the viral capsid antigen VCA-p40.  This antigen is
CC   immunochemically reactive with antibodies to EBV and can be used for
CC   the diagnosis of EBV in a sample.
CC   See also Q51018.
XX
SQ   Sequence 1038 BP; 208 A; 400 C; 278 G; 152 T; 0 other;

  Query Match             5.8%;  Score 37.2;  DB 14;  Length 1038;
  Best Local Similarity   52.6%;  Pred. No. 0.12;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     457 tcttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaactagga 516
           |||||||  ||||  || |  |   ||  | ||||  | |       | | ||    |
Db     468 TCTTTTCCGCTTGTTGGATCGCGAGGGAGCGGGGACGTAGTCTGCCCGCAAGTGGGATCC 409

Qy     517 gcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccct 576
           | ||    ||| || |||  |  || |   || ||||   |||| |   ||| |  ||
Db     408 GTAGTAAGGAGGCACGGGTGGAGGGGGGCCAAAGAGGCCCGGCAAGCCAAAGTAGGCCGG 349

Qy     577 aggtgaggacgtgtgagggtggctgggagaaggg 610
           ||| | |   | ||| ||||||  ||    || |
Db     348 AGGGGGGTGTGGGTGGGGGTGGTAGGTGTGAGCG 315


RESULT   6
A10594/c
ID   A10594 standard; DNA; 10732 BP.
XX
AC   A10594;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Gene encoding a subunit of cellulose synthase.
XX
KW   Cellulose synthase; cellulose production; increase yield; ds.
XX
OS   Vigna angularis.
XX
PN   JP2000060568-A.
XX
PD   29-FEB-2000.
XX
PF   26-AUG-1998;   98JP-0239998.
XX
PR   26-AUG-1998;   98JP-0239998.
XX
PA   (MIZU/) MIZUNO K.
PA   (OJIP ) OJI PAPER CO.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; Y85179.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;


  Query Match             5.7%;  Score 36.6;  DB 21;  Length 10732;
  Best Local Similarity   15.1%;  Pred. No. 0.64;
  Matches   72;  Conservative  190;  Mismatches  210;  Indels    5;  Gaps    1;

Qy       1 ccagaaccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcc 60
           |:::  ::::|::  :   : ::: :|:: ::| :       ||:|:|::: :: |:: :
Db    9779 CSRDSTYSSRGYDANSTSRYSSRYSSTYSSDSTYSAKYCAKSTTBTBCYYDAYDACYDAY 9720

Qy      61 tccacccctccacctatgccatgttttcccttagccactccccagctcggtggaggaaaag 120
           :   :::   : ::   :|| : : : |: ::|:|::  :::: :  :: |:  ::
Db    9719 DANCYSSDSTYTBYCSRRCCCYYDAYSCSRYDARCYDACYSYSNSTCYDATBTTSRYSTT 9660

Qy     121 gcaggcctaactaggtaccgtcttcccgactttgctcaatgatagctgggtgggtctagc 180
           :  : |::| | : | :: : :| :  |  |:::|:     ||:   : ::   ||  :
Db    9659 YSYYSNCYDATTSRCTBYSTBTBTTSRCAKCTBDSTS-----TAKNSTSTYSTTRCTBYS 9605

Qy     181 tgggttccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaac 240
           ::: ::: : :::      |: ::  : | ::|   ::|  :  :::: : :   : :|
Db    9604 SRSRGYSYCSRSRSRCYSCYTDSDSTCYSTTYSTTYAYSCTTSRGYSYDASRSTSTYSRC 9545

Qy     241 tgctacagaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttaa 300
           | :| : :  :: :  ::: ||  :|:    :  :  ::  ::    | |::  :: :
Db    9544 TTSTYSYSTTDYSDCSDYSTTTBNSTYSSDSDCTBYSSDRCSRSDSTCNCYSCSDSRYST 9485

Qy     301 cagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtataggaggca 360
           :    ::|  |:|:|: : |  :|::    : :  ::   :  ||  | | ::   |
Db    9484 TYDACYTYDAAKTBCYTYSDNCCNSTSRCYTNSTYSCSRTBSRNSTCCCTBTTSRGNCCY 9425

Qy     361 agagatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgttcttca 420
           :  :| :|:: : :::      : :| ::::  ::: :| : :: ::  :| :||:::
Db    9424 DAYDANSTSRYDAYYDACYYSYDASTTBYSYSCTBYSNSTYDAYSSRYSCYTCYTCDYSS 9365

Qy     421 tttcaattattcttacctactattcagttcccttgatcttttcttcttgggggctgt 477
           | : :  : |:|   :   :  ||   : |:|:|:: :||: :| :|  ::| | :|
Db    9364 TCYTRCAKCTBCNSTSRAKSRTTTTTTYRCYCYTSRCYTTBCYTCYTNCSRGCCSRT 9308


RESULT   7
V06567
ID   V06567 standard; cDNA; 3614 BP.
XX
AC   V06567;
XX
DT   06-JUL-1998  (first entry)
XX
DE   Arabidopsis cellulose biosynthetic gene clone Ath-B.
XX
KW   Cellulose; cellulose synthase; RSW1 gene; beta-1,4-glucan;
KW   transgenic plant; ss.
XX
OS   Arabidopsis thaliana var. Columbia.
XX
FH   Key             Location/Qualifiers
FT   CDS             217..3411
FT                   /*tag=  a
XX
PN   WO9800549-A1.
XX
PD   08-JAN-1998.
XX
PF   24-JUN-1997;   97WO-AU00402.
```

```
XX
PR   27-JUN-1996;   96AU-0000699.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (AUSU ) UNIV AUSTRALIAN NAT.
XX
PI   Arioli A,  Betzner AS,  Peng L,  Williamson RE;
XX
DR   WPI; 1998-086974/08.
DR   P-PSDB; W33819.
XX
PT   DNA encoding cellulose biosynthetic enzyme - useful for manipulation
PT   of cellulose and beta-1,4-glucan
XX
PS   Claim 12; Page 136-143; 207pp; English.
XX
CC   cDNA clone Ath-B was isolated from an Arabidopsis thaliana cDNA
CC   library using probes obtained by PCR with primers (see T99632-34)
CC   based on cellulose synthase RSW1 genomic clone 23H12 (see V06563)
CC   and EST clone T20782 (see V06562).  It is closely related to RSW1
CC   cDNA (see V06565) and to partial genomic clone 12C4 (see V06564),
CC   and encodes a 1065-polypeptide (see W33819).  Claimed nucleic acid
CC   molecules (see V06562-69) coding for claimed polypeptides (see
CC   W33816-20 and W46202) involved in cellulose biosynthesis can be
CC   used to manipulate the cellulose and/or beta-glucan content of
CC   transgenic plants.  Expression of nucleic acids in the sense
CC   orientation increases the level of cellulose and reduces the level
CC   of non-crystalline beta-1,4-glucan and starch, providing plants
CC   with modified strength and/or shape and/or fibre properties, or
CC   having increased resistance to stresses or pests.  Antisense,
CC   ribozyme or co-suppression molecules can be used to reduce the
CC   cellulose content of a transgenic plant, e.g. to improve
CC   digestibility or to alter carbon partitioning such that increased
CC   carbon is available for growth, rather than deposited as cellulose.
XX
SQ   Sequence 3614 BP; 963 A; 717 C; 863 G; 1071 T; 0 other;
```

```
  Query Match             5.4%;  Score 34.4;  DB 19;  Length 3614;
  Best Local Similarity   52.0%;  Pred. No. 1.8;
  Matches   77;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     340 gaagcatctgtataggaggcaagagatctgagtccttttgaaggcctatcctctgctctg 399
           ||| || ||||  || ||  |  | || |    |  |||  |||  ||       | | |
Db    3401 gaatcaactgttgagaagcgagcaaatatttacctgttttgagggttaaaaaaaacacag 3460

Qy     400 tatctcaattactgttcttcatttcaattattcttacctactattcagttcccttgatct 459
           || | |||||  | ||| |  ||| | || |||     |    ||  |||   |  ||||
Db    3461 aatttaaattatttttcattgttttatttgttcacttttttacttttgttgtgtgtatct 3520

Qy     460 tttcttcttgggggctgtcttagggtca 487
           |   |      |||||| | ||||
Db    3521 gtctgttcgttcttctgtcttggtgtca 3548
```

```
RESULT   8
C76992/c
ID   C76992 standard; cDNA; 1662 BP.
XX
AC   C76992;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2547 polynucleotide sequence SEQ ID NO:5093.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; B42783.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 4281-4282; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1662 BP; 336 A; 516 C; 468 G; 341 T; 1 other;
```

```
  Query Match             5.4%;  Score 34.2;  DB 21;  Length 1662;
  Best Local Similarity   47.1%;  Pred. No. 1.4;
  Matches  105;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy     391 tctgctctgtatctcaattactgttcttcatttcaattattcttacctactattcagttc 450
           | |  | | | |  |    ||  | || ||  |||   || || ||   |  | ||  |
Db    1662 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAG 1603

Qy     451 ccttgatcttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaa 510
               |||   |||  ||  |     ||||||||||||| | ||  ||    ||
Db    1602 TTCATCATCTTTAGACTTAAGGAATTAACAAGGGTCAGGGAGACTACACCAGGCTGAGGG 1543

Qy     511 ctaggagcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaa 570
           ||    |   ||  ||||||||   || |  |||  |   ||| |     | | |  | |
Db    1542 CTTCTTGGTTCCTGGAGACATGCCCAGCTACAGCAAACACAGGGAAACACGAAGGGGGCA 1483
```

```
Qy    571 ttccctaggtgaggacgtgtgagggtggctgggagaagggagg 613
          || |  || | ||     | ||   ||  ||| || |
Db   1482 GCTGGAAGATTTGGTCTTGAACTTGGGGGGTGGGTAAGTGATG 1440
```

RESULT 9
V74828/c
ID V74828 standard; DNA; 4544 BP.
XX
AC V74828;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #517.
XX
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT misc_feature 1..60
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 1801..1860
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3601..3660
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
PN EP786519-A2.
XX
PD 30-JUL-1997.
XX
PF 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
DR WPI; 1997-374922/35.
XX
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
PS Claim 1; Page 1442-1445; 3271pp; English.
XX
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
SQ Sequence 4544 BP; 1717 A; 584 C; 841 G; 1215 T; 187 other;

Query Match 5.3%; Score 33.8; DB 18; Length 4544;
Best Local Similarity 55.6%; Pred. No. 3.1;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy    352 taggaggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaattac 411
          || || ||| | ||  ||  ||   |    |  ||||| |  |||   |||||| ||| |
Db   1549 TACGATGCATGTGACATGCATCTCCTATTTGCTCTATCTTTAGCTTCTTATCTCTATTCC 1490

Qy    412 tgttcttcatttcaattattcttacctactattcagttcccttgatcttttcttctt 468
            | ||| |    ||||| | |    |  | ||||| |  ||| |  | |||   ||
Db   1489 AATACTTTAGCCTAATTACTTTCTTATGATCTTCAGGTAACTTTAAGTATTCACTTT 1433
```

RESULT 10
T62741/c
ID T62741 standard; cDNA to mRNA; 337 BP.
XX
AC T62741;
XX
DT 14-NOV-1997 (first entry)
XX
DE Tobacco acetolactate synthase PCR product.
XX
KW Sulphonylurea herbicide resistance; plant; ALS;
KW polymerase chain reaction; ds.
XX
OS Nicotiana tabacum.
XX
PN WO9708327-A1.
XX
PD 06-MAR-1997.
XX
PF 30-AUG-1996; 96WO-JP02448.
XX
PR 30-AUG-1995; 95JP-0222095.
XX
PA (NISC) NISSAN CHEM IND LTD.
XX
PI Fushimi T, Nakahira K, Nawamaki T, Tagawa M;
XX
DR WPI; 1997-179284/16.
XX
PT Aceto:lactate synthase gene - useful in providing sulphonyl:urea
PT herbicide resistance to plants
XX
PS Example 2; Page 27; 50pp; Japanese.
XX
CC The present sequence represents a PCR product of the amplification
CC of acetolactate synthetase from Tobacco (Nicotiana tabacum). The
CC nucleotide sequence encoding the acetolactate synthase may be used
CC to obtain sulphonylurea herbicide resistant acetolactate synthase
CC protein. The sequence may be used to impart sulphonylurea herbicide
CC resistance to plants by expression in sulphonylurea sensitive plants.
CC The sequence may also be used as a selection marker of plant cells
CC transformed by the nucleic acid.
XX
SQ Sequence 337 BP; 73 A; 144 C; 41 G; 79 T; 0 other;

Query Match 5.3%; Score 33.6; DB 18; Length 337;
Best Local Similarity 57.7%; Pred. No. 0.92;

```
  Matches   60;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     534 ggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgag 593
           || |||||   |||| | ||   ||   || | ||        || | ||   |||| |
Db     215 GGCGTTGGCTGTGAATGTGGGTGGGGGTGAGGTGGAGGGGTGGTGGGGTGGTTTTGTGGG 156

Qy     594 ggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
           |||||  |||| ||||||  ||  | ||  |   ||| || || |
Db     155 GGTGGTGGGGAAAAGGGAAGGTGGATCTAGGGAGGAGGGTGGAG 112


RESULT  11
V55830
ID   V55830 standard; DNA; 795 BP.
XX
AC   V55830;
XX
DT   18-NOV-1998  (first entry)
XX
DE   FLGA insert stabilising polypeptide encoding DNA.
XX
KW   Fusion protein; stabilising polypeptide; proteolytic degradation;
KW   resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW   IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW   nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW   cancer; pathological condition; ss.
XX
OS   Epstein-barr virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..788
FT                   /*tag= a
FT                   /product= "stabilising polypeptide""
XX
PN   WO9822577-A1.
XX
PD   28-MAY-1998.
XX
PF   17-NOV-1997;   97WO-IB01508.
XX
PR   25-JUN-1997;   97US-0048945.
PR   15-NOV-1996;   96US-0030986.
XX
PA   (MASU/) MASUCCI M G.
XX
PI   Masucci MG;
XX
DR   WPI; 1998-312463/27.
DR   P-PSDB; W79128.
XX
PT   New fusion proteins resistant to proteolytic degradation -
PT   comprising a core protein with a stabilising polypeptide comprising
PT   a peptide sequence containing glycine repeats
XX
PS   Disclosure; Fig 3; 120pp; English.
XX
CC   This DNA encodes a stabilising polypeptide and is the FLGA insert of the
CC   invention. The invention provides a method for increasing the resistance
CC   of a core protein to proteolytic degradation that comprises linking or
CC   inserting onto or into the core protein a stabilising polypeptide of
CC   formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC   sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC   Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC   be identical from n repeat to n repeat. Alternatively a nucleic acid
CC   encoding the stabilising polypeptide can be linked onto or inserted into
CC   a nucleic acid encoding a core protein. The fusion proteins of the
CC   invention are more resistant to degradation by proteases and, thus, have
CC   a longer half-life than the unfused core protein. The products can be
CC   used for treating autoimmune diseases, cancer and inflammation. In
CC   particular, the core protein may be an IkappaB regulator protein for the
CC   treatment of inflammatory bowel disease, or a nitroreductase protein
CC   which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC   or other pathological conditions. The fusion proteins can also be used in
CC   diagnostic methods such as in vivo imaging.
XX
SQ   Sequence 795 BP; 200 A; 104 C; 478 G; 13 T; 0 other;


  Query Match             5.3%;  Score 33.6;  DB 19;  Length 795;
  Best Local Similarity   57.7%;  Pred. No. 1.4;
  Matches   60;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     534 ggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgag 593
           ||||  | ||  | |||||||  |||||||  |  | |    |  |   |||| | | |
Db     220 ggaggggcaggagcaggaggaggggcaggaggggggaggggcaggagcaggaggaggggc 279

Qy     594 ggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
           |  || ||  || |||  ||  || || ||   |  |||||||
Db     280 aggagcaggaggaggggcaggaggggcaggagcaggaggaggggg 323


RESULT  12
V89887/c
ID   V89887 standard; cDNA; 269 BP.
XX
AC   V89887;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone CT320.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 362; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 269 BP; 123 A; 49 C; 65 G; 32 T; 0 other;


  Query Match             5.2%;  Score 33.4;  DB 20;  Length 269;
```

```
  Best Local Similarity   54.5%;  Pred. No. 0.95;
  Matches   67;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     363 agatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgttcttcatt 422
           ||  ||| ||| |||||         ||   || |   ||| ||    |||  |||  |
Db     238 AGGGCTGTGTCTTTTTGCTCTAAGTCTCTGCCCTTTTGTTCTAAAGCTTTGTCTTTCTCT 179

Qy     423 tcaattattcttacctactattcagttcccttgatcttttcttcttgggggctgtcttag 482
           ||    ||  |  ||| || |||     |||||   |||||||||  |    ||||||
Db     178 TCTGGAATCTTATCCTTCTGTTCCAGGGCCTTGTCTTTTTCTTCTAAGTCTTTGTCTTTT 119

Qy     483 ggt 485
           |||
Db     118 GGT 116


RESULT  13
C98193/c
ID   C98193 standard; cDNA; 422 BP.
XX
AC   C98193;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:203.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; B53436.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 625; 2104pp; English.
XX
CC   C97991 to C98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in B53234 to B54006. The human
CC   colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. C98764 to C98772 and B54007
CC   represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 422 BP; 96 A; 119 C; 127 G; 78 T; 2 other;


  Query Match             5.2%;  Score 33.4;  DB 21;  Length 422;
  Best Local Similarity   54.5%;  Pred. No. 1.2;
  Matches   67;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     441 tattcagttcccttgatcttttcttcttgggggctgtcttagggtcagggagattgcaga 500
           ||| || | || | | ||||| |||| |||||  |  | |  ||     | |   || |
Db     327 TATCCACTCCCTTGGGTCTTTCCTTCCTGGGGCTTTCCCTGCGGGGCCCGGGGGAGCTGC 268

Qy     501 agcaccagaactaggagcagccctgagacatggggagttggagctgaaggaggaatggca 560
           |||| ||| | | || |  | || |   ||   ||||  | |        | ||| |  |
Db     267 AGCAGCAGCAGTGGGCGACGACCAGCAGCAGGAGCAGCACGGTGCCCATGAAGGTGA 208

Qy     561 gga 563
           |||
Db     207 GGA 205


RESULT  14
A87692/c
ID   A87692 standard; cDNA; 745 BP.
XX
AC   A87692;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Human secreted protein gene 27 SEQ ID NO:37.
XX
KW   Human; secreted protein; immunosuppressive; immunostimulant; nootropic;
KW   antiinflammatory; cardiant; vulnerary; antiulcer; anticonvulsant;
KW   antiparkinsonian; neuroprotective; antiviral; antibacterial; cytostatic;
KW   antiparasitic; thrombolytic; anticoagulant; antiarteriosclerotic;
KW   gene therapy; vaccine; chemotaxis-modulator; angiogenesis-modulator;
KW   cancer; immune system disorder; hyperproliferative disorder; infection;
KW   cardiovascular disorder; neurological disease; wound healing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200043495-A2.
XX
PD   27-JUL-2000.
XX
PF   18-JAN-2000; 2000WO-US00903.
XX
PR   19-JAN-1999;   99US-0116330.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Ebner R,  Young PE,  Ni J,  Moore PA;
PI   Komatsoulis G,  Birse CE;
XX
DR   WPI; 2000-499225/44.
DR   P-PSDB; B25691.
XX
PT   New isolated polynucleotide encoding a secreted protein useful for
PT   preventing, treating or ameliorating a medical condition -
XX
PS   Claim 1; Page 394; 451pp; English.
XX
CC   The polynucleotide sequences given in A87666 to A87708 encodes the human
CC   secreted proteins given in B25665 to B25755. Human secreted proteins
CC   have activities based on the tissues and cells the genes are expressed
CC   in. Examples of activities include: immunosuppressive; immunostimulant;
CC   antiinflammatory; cardiant; vulnerary; antiulcer; nootropic; antiviral;
CC   anticonvulsant; antiparkinsonian; neuroprotective; antibacterial;
CC   antiparasitic; thrombolytic; anticoagulant; antiarteriosclerotic and
CC   cytostatic. The secreted proteins and their polynucleotides can be used
CC   in gene therapy and as vaccines, chemotaxis-modulators and angiogenesis-
CC   modulators. The human secreted proteins and polynucleotides can be used
CC   for diagnosing (the susceptibility to) a pathological condition by
```

CC determining the presence or absence of a mutation in the polynucleotide
CC or determining the presence or amount of expression of the protein. The
CC polynucleotides and proteins can also be used in the treatment and
CC diagnosis of cancer, diseases of the immune system, hyperproliferative
CC disorders, cardiovascular disorders and neurological disease. They can
CC also be used to promote wound healing and to fight infection. A87657 to
CC A87665 and B25664 represent sequences used in the exemplification of the
CC present invention.
XX
SQ Sequence 745 BP; 191 A; 214 C; 188 G; 142 T; 10 other;

Query Match 5.2%; Score 33.4; DB 21; Length 745;
Best Local Similarity 54.5%; Pred. No. 1.6;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy    441 tattcagttcccttgatcttttcttcttgggggctgtcttagggtcagggagattgcaga 500
          ||| || | || | | ||||| |||| |||||  |  | |  ||      | |   || |
Db    621 TATCCACTCCCTTGGGTCTTTCCTTCCTGGGGCTTTCCCTGCGGGGCCCGGGGGAGCTGC 562

Qy    501 agcaccagaactaggagcagccctgagacatggggagttggagctgaaggaggaatggca 560
          |||| ||| | | || |  | || |   ||   ||||  | |         | ||| |  |
Db    561 AGCAGCAGCAGTGGGCGACGACCAGCAGCAGCAGGAGCAGCACGGTGCCCATGAAGGTGA 502

Qy    561 gga 563
          |||
Db    501 GGA 499
```

RESULT 15
Z43768/c
ID Z43768 standard; cDNA; 9312 BP.
XX
AC Z43768;
XX
DT 08-MAR-2000 (first entry)
XX
DE Human CRP-ductin cDNA.
XX
KW CRP-ductin; human; SCUZ protein;;cytostatic; treatment; carcinoma;
KW brain tumor; breast tumor; lung; melanoma; pancreas; endometrium;
KW oesophagus; colon; stomach; ss.
XX
OS Homo sapiens.
XX
PN DE19829660-C1.
XX
PD 23-DEC-1999.
XX
PF 02-JUL-1998; 98DE-1029660.
XX
PR 02-JUL-1998; 98DE-1029660.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Mollenhauer J, Poustka A, Krebs I;
XX
DR WPI; 2000-054563/05.
XX
PT Transgenic non-human mammal useful for testing compounds for anticancer
PT activity -
XX
PS Claim 14; Fig 1; 14pp; German.
XX
CC This invention describes a novel non-human mammal, comprises a gene,
CC that codes for a SCUZ protein, that is altered so that it is switched
CC off, or can be switched off in a time- or tissue-specific manner. The
CC protein described in the invention has cytostatic activity. The non-human
CC mammal or cells derived from it are useful for testing of compounds that
CC influence the function, in particular the activity of a SCUZ protein. The
CC compounds are useful for treatment of carcinoma, in particular brain or
CC breast tumors, lung carcinoma, melanoma or pancreatic, endometrial,
CC oesophageal, colon or stomach carcinomas. This sequence represents a
CC nucleic acid fragment which encodes the human CRP-ductin gene exon 1 and
CC its promoter which is used in the method of the invention.
XX
SQ Sequence 9312 BP; 2266 A; 2240 C; 2140 G; 2664 T; 2 other;

Query Match 5.2%; Score 33.4; DB 21; Length 9312;
Best Local Similarity 52.5%; Pred. No. 5.9;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy    487 agggagattgcagaagcaccagaactaggagcagccctgagacatggggagttggagctg 546
          | ||||  |  ||  |   ||  || || ||     || | |  |||  | ||  |
Db    342 AAGGAGGAGGAGGAGAGAAGGAAAAGAGAAGGAGAAGGAGGAGAAGAAGAGAAGAAGAAG 283

Qy    547 aaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtggctgggaga 606
          ||||||||  | |||| || ||       |  | ||  | | |||||  |  ||||||
Db    282 AAGGAGGAGGAGGAGGAGGAGGAGGGGAGGGGAGGGGGGAGGGGGAGGGGGAGAGGGAGA 223

Qy    607 agggaggggtggtcacgaa 625
          ||  | | ||  | |||
Db    222 GGGAGAGAGGGGGAAGGAA 204
```

Search completed: April 30, 2001, 19:07:21
Job time: 36855 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:33:02 ; Search time 330.07 Seconds
(without alignments)
338.023 Million cell updates/sec

Title: US-09-721-183-4
Perfect score: 639
Sequence: 1 ccagaaccgagtttaggtcc.....cacgaatggacggaggggat 639

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	45.2	7.1	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	42	6.6	53526	3	US-08-658-136-2	Sequence 2, Appli
c	3	42	6.6	53577	3	US-08-658-136-1	Sequence 1, Appli
	4	39.8	6.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c	5	37.2	5.8	1038	1	US-08-031-148-3	Sequence 3, Appli
c	6	37.2	5.8	1038	3	US-08-415-838-3	Sequence 3, Appli
c	7	32.4	5.1	3680	1	US-08-759-848-1	Sequence 1, Appli
c	8	32.4	5.1	3680	5	PCT-US95-09383-1	Sequence 1, Appli
	9	31.8	5.0	1555	2	US-08-696-376-1	Sequence 1, Appli
c	10	30.8	4.8	1643	2	US-08-933-750C-68	Sequence 68, Appl
c	11	30.8	4.8	1643	4	US-09-234-613-68	Sequence 68, Appl
c	12	30.4	4.8	2862	1	US-08-148-209A-1	Sequence 1, Appli
c	13	30.2	4.7	12847	1	US-08-550-715-1	Sequence 1, Appli
	14	29.8	4.7	1866	3	US-09-173-581-13	Sequence 13, Appl
	15	29.8	4.7	4095	5	PCT-US91-09422-18	Sequence 18, Appl
	16	29.8	4.7	53526	3	US-08-658-136-2	Sequence 2, Appli
	17	29.8	4.7	53577	3	US-08-658-136-1	Sequence 1, Appli
c	18	29.4	4.6	390	2	US-08-429-764A-4	Sequence 4, Appli
	19	29.4	4.6	390	2	US-08-429-764A-5	Sequence 5, Appli
c	20	29.4	4.6	4823	2	US-08-457-254-5	Sequence 5, Appli
c	21	29.4	4.6	4823	2	US-08-484-257-20	Sequence 20, Appl
c	22	29.4	4.6	4823	3	US-08-999-927-5	Sequence 5, Appli
c	23	29.4	4.6	4823	5	PCT-US94-08806-28	Sequence 28, Appl
c	24	29.4	4.6	4823	5	PCT-US95-01829-5	Sequence 5, Appli
c	25	29.4	4.6	4823	5	PCT-US95-16626-5	Sequence 5, Appli
	26	29.2	4.6	325	2	US-08-332-766A-11	Sequence 11, Appl
c	27	29.2	4.6	554	3	US-08-840-146-16	Sequence 16, Appl
c	28	29.2	4.6	554	3	US-09-360-220-16	Sequence 16, Appl
c	29	29.2	4.6	2241	3	US-08-840-146-1	Sequence 1, Appli
c	30	29.2	4.6	2241	3	US-09-360-220-1	Sequence 1, Appli
	31	29.2	4.6	5526	4	US-08-751-359-21	Sequence 21, Appl
c	32	29	4.5	10684	3	US-08-618-100B-3	Sequence 3, Appli
	33	28.8	4.5	7868	6	5175383-1	Patent No. 5175383
	34	28.6	4.5	659	2	US-08-454-115-1	Sequence 1, Appli
	35	28.6	4.5	1242	2	US-08-454-115-4	Sequence 4, Appli
c	36	28.6	4.5	1371	3	US-08-462-969B-1	Sequence 1, Appli
	37	28.6	4.5	2427	1	US-08-490-099-1	Sequence 1, Appli
	38	28.6	4.5	2675	1	US-08-232-079-1	Sequence 1, Appli
	39	28.6	4.5	3337	1	US-08-072-610-1	Sequence 1, Appli
	40	28.6	4.5	3337	2	US-08-719-822B-1	Sequence 1, Appli
	41	28.6	4.5	4147	3	US-08-560-005-1	Sequence 1, Appli
	42	28.6	4.5	11613	1	US-08-484-044-10	Sequence 10, Appl
	43	28.4	4.4	1386	2	US-08-408-095-22	Sequence 22, Appl
	44	28.4	4.4	3708	1	US-08-185-232A-1	Sequence 1, Appli
	45	28.4	4.4	3708	1	US-08-416-523-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             7.1%;  Score 45.2;  DB 1;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 0.00033;
  Matches    5;  Conservative  111;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     479 ttagggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagt 538
           || ||  |::::::::  : ::::: :  ::::  ::::: ::    :::: : ::::::
Db    1442 TTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1383

Qy     539 tggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtgg 598
           ::::  :::::::::: :: :::: ::::::      ::: ::::: : : ::::: ::
Db    1382 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1323

Qy     599 ctgggagaagggaggggtggtcacgaatggacggagggga 638
           ::::::::::::::: ::  : ::: ::: ::::::::
Db    1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1283


RESULT   2
US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-2


  Query Match             6.6%;  Score 42;  DB 3;  Length 53526;
  Best Local Similarity   52.9%;  Pred. No. 0.0091;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db   34695 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGGGGAGGAGGGAGGAGG 34636

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
```

```
           | | || |  ||||  |  ||  | | || |||| | || |        | | ||||
Db   34635 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGGAGGAGGGGGGACAG 34576

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
           | | ||  ||   ||||| ||||||   ||  | ||  ||| || | |||
Db   34575 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34526


RESULT   3
US-08-658-136-1/c
; Sequence 1, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53577 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-1


  Query Match             6.6%;  Score 42;  DB 3;  Length 53577;
  Best Local Similarity   52.9%;  Pred. No. 0.0091;
  Matches   90;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           ||||   |    ||||   | | ||     | || |  || |   |||| ||    |||
Db   34690 GGGGAGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGAGGAGGGGGAGGAGGGAGGAGG 34631

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
           | | || |  ||||  |  ||  | | || |||| | || |        | | ||||
Db   34630 GAGGAGGGGGAGGAGGAGGGGGGAGGAGGGGAGGAGGGAGGAGGGGAGGAGGGGGGACAG 34571

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
           | | ||  ||   ||||| ||||||   ||  | ||  ||| || | |||
Db   34570 GAGGGGAGGGGGAGGAGAGGGGAGGAAGGGGGAGGAGGGGAGGGGGAGGA 34521


RESULT   4
```

```
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             6.2%;  Score 39.8;  DB 1;  Length 7218;
  Best Local Similarity   5.4%;  Pred. No. 0.016;
  Matches   17;  Conservative  167;  Mismatches  129;  Indels    0;  Gaps    0;

Qy       7 ccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcctccacc 66
           ||||| || | | |||||    |         ||:::::::::: :::::::::::: ::
Db    1033 CCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYY 1092

Qy      67 cctccacctatgccatgttttcccttagccactccccagctcggtggaggaaaggcaggc 126
           ::::: ::: : :: : ::::::::::  :: ::::::  :::  :          :   :
Db    1093 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1152

Qy     127 ctaactaggtaccgtcttcccgactttgctcaatgatagctgggtgggtctagctgggtt 186
           ::  ::   : :: :::::::  :::: :::  :  :  ::   :   :::  ::   ::
Db    1153 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1212

Qy     187 ccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaactgctac 246
           ::  :: ::: :  : :     : :::::: ::::  :::: :   :     :  :: :: :
Db    1213 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1272

Qy     247 agaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttaacagcag 306
             :   : :  ::  :::: : :::: :::::: ::: :::::   :::  :  :
Db    1273 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1332

Qy     307 gtaatctgcttct 319
           :  ::: :::::
Db    1333 YYYYYYYYYYYYY 1345


RESULT   5
US-08-031-148-3/c
; Sequence 3, Application US/08031148
; Patent No. 5424398
;  GENERAL INFORMATION:
;    APPLICANT:  Middeldorp, Jaap Michiel.
;    TITLE OF INVENTION:  Peptides and nucleic acid sequences
;    TITLE OF INVENTION:  related to the Epstein-Barr virus.
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo Pharma
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850-4377
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/031,148
;      FILING DATE:  19930312
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92200721.6
;      FILING DATE:  13-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bobrowicz, Donna
;      REGISTRATION NUMBER:  32,196
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1038 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Epstein-Barr virus
US-08-031-148-3


  Query Match             5.8%;  Score 37.2;  DB 1;  Length 1038;
  Best Local Similarity   52.6%;  Pred. No. 0.038;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     457 tctttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaactagga 516
           |||||||  ||||  || |  |   ||  | ||||  | |      | | ||    |
Db     468 TCTTTTCCGCTTGTTGGATCGCGAGGGAGCGGGGACGTAGTCTGCCCGCAAGTGGGATCC 409

Qy     517 gcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccct 576
           | ||    ||| || |||  |  || |   || ||||   ||||  |   ||| |  ||
Db     408 GTAGTAAGGAGGCACGGGTGGAGGGGGGCCAAAGAGGCCCGGCAAGCCAAAGTAGGCCGG 349

Qy     577 aggtgaggacgtgtgagggtggctgggagaaggg 610
           ||| | |   | ||| ||||||  ||    || |
Db     348 AGGGGGGTGTGGGTGGGGGTGGTAGGTGTGAGCG 315


RESULT   6
US-08-415-838-3/c
; Sequence 3, Application US/08415838
; Patent No. 6008327
;  GENERAL INFORMATION:
;    APPLICANT:  Middeldorp, Jaap Michiel.
;    TITLE OF INVENTION:  Peptides and nucleic acid sequences
;    TITLE OF INVENTION:  related to the Epstein-Barr virus.
```

```
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo-No. 6008327el Patent Department
;      STREET:  1300 Piccard Drive, Suite 206
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,838
;      FILING DATE:  03-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92200721.6
;      FILING DATE:  13-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gormley, Mary E.
;      REGISTRATION NUMBER:  34,409
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1038 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Epstein-Barr virus
US-08-415-838-3


  Query Match             5.8%;  Score 37.2;  DB 3;  Length 1038;
  Best Local Similarity   52.6%;  Pred. No. 0.038;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy     457 tcttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaactagga 516
           ||||||| ||||  || | |   ||  | ||||  | |       | | ||    |
Db     468 TCTTTTCCGCTTGTTGGATCGCGAGGGAGCGGGGACGTAGTCTGCCCGCAAGTGGGATCC 409

Qy     517 gcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccct 576
           | ||    ||| || |||  |  || | . || ||||   |||| |   ||| |  ||
Db     408 GTAGTAAGGAGGCACGGGTGGAGGGGGGCCAAAGAGGCCCGGCAAGCCAAAGTAGGCCGG 349

Qy     577 aggtgaggacgtgtgagggtggctgggagaaggg 610
           ||| | |   | ||| ||||||  ||    || |
Db     348 AGGGGGGTGTGGGTGGGGGTGGTAGGTGTGAGCG 315


RESULT   7
US-08-759-848-1/c
; Sequence 1, Application US/08759848
; Patent No. 5750826
;  GENERAL INFORMATION:
;    APPLICANT:  Borkowski, Joseph A.
;    APPLICANT:  Strader, Catherine D.
;    APPLICANT:  Hess, John W.
;    APPLICANT:  Chen, Howard Y.
;    APPLICANT:  Trumbauer, Myrna E.
;    TITLE OF INVENTION:  BRADYKININ B2 RECEPTOR MODIFIED
;    TITLE OF INVENTION:  NON-HUMAN ANIMALS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John W. Wallen III
;      STREET:  126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065-0900
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/759,848
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/281,393
;      FILING DATE:  27-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wallen III, John W.
;      REGISTRATION NUMBER:  35,403
;      REFERENCE/DOCKET NUMBER:  19234
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3905
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3680 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-759-848-1


  Query Match             5.1%;  Score 32.4;  DB 1;  Length 3680;
  Best Local Similarity   49.4%;  Pred. No. 2.2;
  Matches   84;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           | |||  |    || |  |  | ||| |||||||  |  ||    |||  |||    |||
Db    1258 GAGGGGAGGGGAAGAGAAAACGGGATAGCAGAAGGAGGAGGTGGAGGGCCAGGGAGGAGC 1199

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
             ||  || |||| | ||| |   | ||    | ||| || | ||   ||    |  ||
Db    1198 ACAGGACAGGTGGACCAGAAAGTCTATTGTTCAGCTGATGACTCCCGGCGGCAGTGGGGT 1139

Qy     589 gtgagggtggctgggagaagggagggggtggtcacgaatggacggagggga 638
           |||  ||   |    ||       |    |||| ||  | | |  |    |||
Db    1138 GTGGGGAGGTGGTGGGACTCCGGTCTGTGGACAGCACTAGGGGACTTGGA 1089


RESULT   8
PCT-US95-09383-1/c
; Sequence 1, Application PC/TUS9509383
;  GENERAL INFORMATION:
;    APPLICANT:  Borkowski, Joseph A.
;    APPLICANT:  Strader, Catherine D.
;    APPLICANT:  Hess, John W.
;    APPLICANT:  Chen, Howard Y.
;    APPLICANT:  Trumbauer, Myrna E.
;    TITLE OF INVENTION:  BRADYKININ B2 RECEPTOR MODIFIED
;    TITLE OF INVENTION:  NON-HUMAN ANIMALS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John W. Wallen III
;      STREET:  126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065-0900
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09383
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/281,393
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wallen III, John W.
;      REGISTRATION NUMBER:  35,403
;      REFERENCE/DOCKET NUMBER:  19234
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3905
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3680 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US95-09383-1


  Query Match             5.1%;  Score 32.4;  DB 5;  Length 3680;
  Best Local Similarity   49.4%;  Pred. No. 2.2;
  Matches   84;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           | |||  |    || |  |  | ||| |||||||  |  ||   |||  |||    |||
Db    1258 GAGGGGAGGGGAAGAGAAAACGGGATAGCAGAAGGAGGAGGTGGAGGGCCAGGGAGGAGC 1199

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
             ||  || |||| | ||| |   | ||    | ||| || | ||   ||    |  ||
Db    1198 ACAGGACAGGTGGACCAGAAAGTCTATTGTTCAGCTGATGACTCCCGGCGGCAGTGGGGT 1139

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
           |||  ||  |    ||      |    |||| ||  | | |  |    |||
Db    1138 GTGGGGAGGTGGTGGGACTCCGGTCTGTGGACAGCACTAGGGGACTTGGA 1089


RESULT  9
US-08-696-376-1
; Sequence 1, Application US/08696376
; Patent No. 5830723
;  GENERAL INFORMATION:
;    APPLICANT:  FOSTER, DOUGLAS N.
;    APPLICANT:  FARRIS, JAMES A.
;    APPLICANT:  FOSTER, LINDA K.
;    TITLE OF INVENTION:  METHOD FOR IMMORTALIZING CELLS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
;      STREET:  119 NORTH FOURTH STREET, #203
;      CITY:  MINNEAPOLIS
;      STATE:  MINNESOTA
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,376
;      FILING DATE:  13-AUG-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MCCORMACK, MYRA H.
;      REGISTRATION NUMBER:  36,602
;      REFERENCE/DOCKET NUMBER:  110.00400101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1225
;      TELEFAX:  612-305-1228
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1555 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-696-376-1


  Query Match             5.0%;  Score 31.8;  DB 2;  Length 1555;
  Best Local Similarity   67.2%;  Pred. No. 2.2;
  Matches   45;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      42 tttctccttaccttcttcctccacccctccacctatgccatgttttcccttagccactcc 101
           |||| ||||  |  |  || |   |||| || | | |||   ||| ||||||||||||||
Db    1263 TTTCCCCTTTACCACGCCCCCTTTCCCTTCAGCCACGCCCCCTTTCCCCTTAGCCACTCC 1322

Qy     102 ccagctc 108
           ||  | |
Db    1323 CCTTCCC 1329


RESULT  10
US-08-933-750C-68/c
; Sequence 68, Application US/08933750C
; Patent No. 5932442
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750C
;      FILING DATE:  September 23, 1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1643 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SYNOOAT01
;      CLONE:  724157
US-08-933-750C-68
```

```
  Query Match             4.8%;  Score 30.8;  DB 2;  Length 1643;
  Best Local Similarity   50.0%;  Pred. No. 4.7;
  Matches   77;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     460 tttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaactaggagca 519
           |||   |||  ||  |     |||||||||||||| | ||  ||     ||  ||     |
Db    1620 TTTAGACTTAAGGAATTAACAAGGGTCAGGGAGACTACACCAGGCTGAGGGCTTCTTGGT 1561

Qy     520 gccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccctagg 579
           ||  ||||||||   || |  |||  |   ||| |     | | |  | |      ||
Db    1560 TCCTGGAGACATGCCCAGCTACAGCAAACACAGGGAAACACGAAGGGGGCAGCTGGAAGA 1501

Qy     580 tgaggacgtgtgagggtggctgggagaagggagg 613
           |  || | ||     | ||   ||  ||| || |
Db    1500 TTTGGTCTTGAACTTGGGGGGTGGGTAAGTGATG 1467


RESULT  11
US-09-234-613-68/c
; Sequence 68, Application US/09234613
; Patent No. 6132973
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/234,613
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750
;      FILING DATE:  September 23, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  68:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1643 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SYNOOAT01
;      CLONE:  724157
US-09-234-613-68


  Query Match             4.8%;  Score 30.8;  DB 4;  Length 1643;
  Best Local Similarity   50.0%;  Pred. No. 4.7;
  Matches   77;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     460 tttcttcttgggggctgtcttagggtcagggagattgcagaagcaccagaactaggagca 519
           |||   |||  ||  |     |||||||||||||| | ||  ||     ||  ||     |
Db    1620 TTTAGACTTAAGGAATTAACAAGGGTCAGGGAGACTACACCAGGCTGAGGGCTTCTTGGT 1561

Qy     520 gccctgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccctagg 579
           ||  ||||||||   || |  |||  |   ||| |     | | |  | |      ||
Db    1560 TCCTGGAGACATGCCCAGCTACAGCAAACACAGGGAAACACGAAGGGGGCAGCTGGAAGA 1501

Qy     580 tgaggacgtgtgagggtggctgggagaagggagg 613
           |  || | ||     | ||   ||  ||| || |
Db    1500 TTTGGTCTTGAACTTGGGGGGTGGGTAAGTGATG 1467


RESULT  12
US-08-148-209A-1/c
; Sequence 1, Application US/08148209A
; Patent No. 5556780
;  GENERAL INFORMATION:
;    APPLICANT:  Dzau, Victor J
;    APPLICANT:  Mukoyama, Masashi
;    TITLE OF INVENTION:  TYPE-2 ANGIOTENSIN II RECEPTOR AND GENE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/148,209A
;      FILING DATE:  05-NOV-1993
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20,015
;      REFERENCE/DOCKET NUMBER:  A-58491-1/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2862 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  132..1223
US-08-148-209A-1


  Query Match             4.8%;  Score 30.4;  DB 1;  Length 2862;
  Best Local Similarity   50.7%;  Pred. No. 8.3;
  Matches   73;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     296 cttaacagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtatagg 355
           | ||| |    ||  | |  || ||   |    | |   |||||||     | | | |
Db    2047 CATAATACAGTGTTTTGTTGTTGTATTCAAAAATCATCAGAAGAGAATAGAATAAACAAG 1988

Qy     356 aggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgtt 415
           || ||| |    |||    |||| || |  |||  || | | |||    |     | | |
Db    1987 AGTCAAAATGGATGAAAAATTTTCAAAGTGTATATTCAGTTTTGTGGATACTAAAATACT 1928
```

```
Qy    416 cttcatttcaattattcttaccta 439
          |||||| || || |  ||  |||
Db   1927 TTTCATTGCAGTTCTGTTTTACTA 1904


RESULT 13
US-08-550-715-1/c
; Sequence 1, Application US/08550715
; Patent No. 5750345
;  GENERAL INFORMATION:
;    APPLICANT:  Bowie, Lemuel J.
;    TITLE OF INVENTION:  Human `-Thalassemia Mutations as a Predictor of
;    TITLE OF INVENTION:  Blood-Related Disorders
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/550,715
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28493/32834
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12847 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(6703..6797, 6915..7119, 7262..7387)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(10514..10608, 10726..10930, 11080..11205)
US-08-550-715-1


  Query Match             4.7%;  Score 30.2;  DB 1;  Length 12847;
  Best Local Similarity   53.9%;  Pred. No. 21;
  Matches   62;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    524 tgagacatggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgag 583
          |||||| |  |||  || |   |  ||||||   | |||| | || |  |       | |
Db   9418 TGAGACTTAAGGATATGTATTAGGTGGAGGAGGTGGAGGAGGGAGGAGGCTTGGCTCGGG 9359

Qy    584 gacgtgtgagggtggctgggagaagggaggggtggtcacgaatggacggagggga 638
          || | |  |||   |  | || |||||    |||||     |  |  || |||||
Db   9358 GAGGGCAAAGGACAGAGGAGACAAGGGCCCTGTGGTTGGAGAATGGAGGTGGGGA 9304


RESULT 14
US-09-173-581-13
; Sequence 13, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/173,581A
;  CURRENT FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 13
;   LENGTH: 1866
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 1439938
US-09-173-581-13


  Query Match             4.7%;  Score 29.8;  DB 3;  Length 1866;
  Best Local Similarity   55.2%;  Pred. No. 10;
  Matches   58;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    535 gagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagg 594
          |||  | ||  || | ||  | ||       || ||       || |||||  | ||| |
Db   1303 gagaggaagtggaggcagttaggggcatggaggagagagagagatgagggagagtggga 1362

Qy    595 gtggctgggagaagggaggggtggtcacgaatggacggaggggat 639
          |     ||  || ||  |||||||  | |||  || |||||  |||
Db   1363 gcacggggcggatggacggggtggagaagaagagagggaggagat 1407


RESULT 15
PCT-US91-09422-18
; Sequence 18, Application PC/TUS9109422
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R.
;    APPLICANT:  Hagen, Frederick S.
;    APPLICANT:  Houamed, Khaled M.
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN-COUPLED GLUTAMATE RECEPTORS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09422
;      FILING DATE:  19911212
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
```

```
;       REGISTRATION NUMBER:  31,990
;       REFERENCE/DOCKET NUMBER:  13952-6PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 467-9600
;       TELEFAX:  (415) 543-5043
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4095 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  SN30
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  463..3198
PCT-US91-09422-18


  Query Match             4.7%;  Score 29.8;  DB 5;  Length 4095;
  Best Local Similarity   51.9%;  Pred. No. 15;
  Matches   67;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     289 tcccaaacttaacagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatct 348
           |||| | |   | || ||   |||| | | |  ||   ||  |  | ||     | ||
Db     149 TCCCTACCAGGATAGGAGCCTATCTCCCTTTCACAGCAGGACACAGAAATCTGGCCTTCA 208

Qy     349 gtataggaggcaagagatctgagtccttttgaaggcctatcctctgctctgtatctcaat 408
           |||     || ||  |||||||| |||  || ||  ||  || |||      ||||
Db     209 GTACTTTGGGAAAAGGATCTGAGACCTCCTGGAGCTCTGACCACTGGCTGTCATCTGTGG 268

Qy     409 tactgttct 417
             |||  ||
Db     269 CTCTGGCCT 277


Search completed: May  1, 2001, 00:33:32
Job time: 56556 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:17:41 ; Search time 12176.5 Seconds
(without alignments)
7.670 Million cell updates/sec

Title: US-09-721-183-4
Perfect score: 639
Sequence: 1 ccagaaccgagtttaggtcc.....cacgaatggacggaggggat 639

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	545.4	85.4	558	205	AQ343789	AQ343789 RPCI11-11
2	403	63.1	416	1	AA026755	AA026755 ze93a12.r
c 3	387	60.6	388	202	AQ134191	AQ134191 HS_3057_A
c 4	296.4	46.4	369	146	BF330839	BF330839 RC3-BT033
5	97.8	15.3	229	120	AW806456	AW806456 MR4-UM011
c 6	62.8	9.8	148	120	AW806459	AW806459 MR4-UM011
7	62.2	9.7	473	165	BE284691	BE284691 601098104
8	60.6	9.5	519	142	BF021610	BF021610 uy49f05.y
c 9	55.8	8.7	411	162	BE062071	BE062071 RC1-BT025
10	44.2	6.9	1451	172	BG066225	BG066225 H3040D10-
11	43.4	6.8	1201	229	CNS016AN	AL106505 Drosophil
c 12	41.6	6.5	869	227	B21587	B21587 T29I6-Sp6 T
13	41.6	6.5	872	121	AW926672	AW926672 HVSMEg000
14	41.6	6.5	1097	145	BF256465	BF256465 HVSMEf001
15	41.2	6.4	307	105	AL513839	AL513839 AL513839
16	41	6.4	884	229	CNS0129A	AL101272 Drosophil
17	40.6	6.4	1011	210	AQ738680	AQ738680 HS_5382_B
18	40.4	6.3	802	229	CNS011T0	AL100686 Drosophil

c	19	40.4	6.3	874	227	B20834	B20834 T28C4-T7 TA
c	20	40.2	6.3	428	227	B35843	B35843 HS-1031-A2-
c	21	40.2	6.3	987	229	CNS00418	AL066537 Drosophil
c	22	40.2	6.3	1101	229	CNS00L4E	AL067752 Drosophil
c	23	40	6.3	939	229	CNS00CNG	AL059400 Drosophil
c	24	40	6.3	1407	227	B09311	B09311 T25B16-T7 T
c	25	40	6.3	1581	173	BG120008	BG120008 602352294
	26	39.8	6.2	523	173	BG109361	BG109361 602280414
	27	39.6	6.2	328	105	AL515163	AL515163 AL515163
	28	39.6	6.2	1280	145	BF256536	BF256536 HVSMEf001
	29	39.4	6.2	951	225	AZ672893	AZ672893 ENTJT74TF
	30	39.2	6.1	592	117	AW585213	AW585213 N211487e
c	31	39.2	6.1	838	227	B19873	B19873 T18C3-Sp6 T
	32	38.8	6.1	994	170	BF864880	BF864880 963055F04
	33	38.8	6.1	1291	170	BF862959	BF862959 963039H08
c	34	38.6	6.0	489	216	AZ128901	AZ128901 OSJNBb006
c	35	38.6	6.0	994	231	CNS04187	AL269872 Tetraodon
	36	38.6	6.0	1101	229	CNS0181N	AL108773 Drosophil
	37	38.6	6.0	1151	145	BF256631	BF256631 HVSMEf001
	38	38.4	6.0	1239	145	BF256577	BF256577 HVSMEf001
	39	38.2	6.0	373	105	AL514185	AL514185 AL514185
	40	38.2	6.0	1077	229	CNS0156Q	AL105068 Drosophil
c	41	38	5.9	784	227	B19307	B19307 F1F22-T7 IG
c	42	38	5.9	860	231	CNS041Q7	AL270520 Tetraodon
c	43	38	5.9	1014	146	BF262846	BF262846 HV_CEa000
c	44	38	5.9	1201	229	CNS00ZON	AL097937 Drosophil
	45	37.8	5.9	1107	170	BF867881	BF867881 963094E07

ALIGNMENTS

```
RESULT  1
AQ343789/c
LOCUS       AQ343789      558 bp    DNA             GSS       07-MAY-1999
DEFINITION  RPCI11-115I7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-115I7,
            DNA sequence.
ACCESSION   AQ343789
VERSION     AQ343789.1  GI:4168685
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-115I7.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Homo sapiens"
                     /db_xref="GDB:7543974"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-115I7"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      165 a    130 c    136 g    127 t
ORIGIN


  Query Match             85.4%;  Score 545.4;  DB 205;  Length 558;
  Best Local Similarity   99.6%;  Pred. No. 3.3e-145;
  Matches  557;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      15 aggtccaggttctcgttctggcaaatcttttcttccttaccttcttcctccacccctccacc 74
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     558 AGGTCCAGGTTCTCGTTCTGGC-AATCTTTCTCCTTACCTTCTTCCTCCACCCCTCCACT 500

Qy      75 tatgccatgttttcccttagccactccccagctcggtggaggaaaggcaggcctaactag 134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TATGCCATGTTTTCCCTTAGCCACTCCCCAGCTCGGTGGAGGAAAGGCAGGCCTAACTAG 440

Qy     135 gtaccgtcttcccgactttgctcaatgatagctgggtgggtctagctgggttccagccac 194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 GTACCGTCTTCCCGACTTTGCTCAATGATAGCTGGGTGGGTCTAGCTGGGTTCCAGCCAC 380

Qy     195 ttgtaatgtgggacatctctcaccccaactttgtaggtggagcaactgctacagaggtaa 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 TTGTAATGTGGGACATCTCTCACCCCAACTTTGTAGGTGGAGCAACTGCTACAGAGGTAA 320

Qy     255 atatgattaactttacattccatctttcgtctgctcccaaacttaacagcaggtaatctg 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ATATGATTAACTTTACATTCCATCTTTCGTCTGCTCCCAAACTTAACAGCAGGTAATCTG 260

Qy     315 cttctagcaagtggtgaaggtaagagaagcatctgtataggaggcaagagatctgagtcc 374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 CTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATCTGTATAGGAGGCAAGAGATCTGAGTCC 200

Qy     375 ttttgaaggcctatcctctgctctgtatctcaattactgttcttcatttcaattattctt 434
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 TTTTGAAGGCCTATCCTCTGCTCTGTATCTCAATTACTGTTCTTCATTTCAATTATTCTT 140

Qy     435 acctactattcagttcccttgatctttttcttcttgggggctgtcttagggtcagggagat 494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 ACCTACTATTCAGTTCCCTTGATCTTTTCTTCTTGGGGGCTGTCTTAGGGTCAGGGAGAT 80

Qy     495 tgcagaagcaccagaactaggagcagccctgagacatggggagttggagctgaaggagga 554
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 TGCAGAAGCACCAGAACTAGGAGCAGCCCTGAGACATGGGGAGTTGGAGCTGAAGGAGGA 20

Qy     555 atggcaggatgaagaattc 573
           ||||||||||||||||||
Db      19 ATGGCAGGATGAAGAATTC 1


RESULT  2
AA026755
LOCUS       AA026755      416 bp    mRNA            EST       01-FEB-1997
DEFINITION  ze93a12.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:366526 5', mRNA sequence.
ACCESSION   AA026755
VERSION     AA026755.1  GI:1492553
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
```

```
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1155   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 386.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Homo sapiens"
                     /db_xref="GDB:1283230"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:366526"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT       93 a    111 c     89 g    122 t      1 others
ORIGIN


  Query Match             63.1%;  Score 403;  DB 1;  Length 416;
  Best Local Similarity   99.8%;  Pred. No. 1.6e-104;
  Matches  414;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 ccagaaccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CCAGAACCGAGTTTAGGTCCAGGTTCTCGTTCTGGCAAATCTTTCTCCTTACCTTCTTCC 61

Qy     61 tccacccctccacctatgccatgttttcccttagccactccccagctcggtggaggaaag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TCCACCCCTCCACCTATGCCATGTTTTCCCTTAGCCACTCCCCAGCTCGGTGGAGGAAAG 121

Qy    121 gcaggcctaactaggtaccgtcttcccgactttgctcaatgatagctgggtgggtctagc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GCAGGCCTAACTAGGTACCGTCTTCCCGACTTTGCTCAATGATAGCTGGGTGGGTCTAGC 181

Qy    181 tgggttccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 TGGGTTCCAGCCACTTGTAATGTGGGACATCTCTCACCCCAACTTTGTAGGTGGAGCAAC 241

Qy    241 tgctacagaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 TGCTACAGAGGTAAATATGATTAACTTTACATTCCATCTTTCGTCTGCTCCCAAACTTAA 301

Qy    301 cagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtataggaggca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 CAGCAGGTAATCTGCTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATCTGTATAGGAGGCA 361

Qy    361 agagatc-tgagtccttttgaaggcctatcctctgctctgtatctcaattactgt 414
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AGAGATCTTGAGTCCTTTTGAAGGCCTATCCTCTGCTCTGTATCTCAATTACTGT 416


RESULT   3
```

```
AQ134191/c
LOCUS       AQ134191      388 bp    DNA             GSS       23-SEP-1998
DEFINITION  HS_3057_A2_G05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3057 Col=10 Row=M, DNA sequence.
ACCESSION   AQ134191
VERSION     AQ134191.1  GI:3525557
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3057  row: M  column: 10
            Class: BAC ends
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3057 Col=10 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      114 a     95 c     81 g     98 t
ORIGIN


  Query Match             60.6%;  Score 387;  DB 202;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 6e-100;
  Matches  387;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    181 tgggttccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 TGGGTTCCAGCCACTTGTAATGTGGGACATCTCTCACCCCAACTTTGTAGGTGGAGCAAC 328

Qy    241 tgctacagaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 TGCTACAGAGGTAAATATGATTAACTTTACATTCCATCTTTCGTCTGCTCCCAAACTTAA 268

Qy    301 cagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtataggaggca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 CAGCAGGTAATCTGCTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATCTGTATAGGAGGCA 208

Qy    361 agagatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgttcttca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 AGAGATCTGAGTCCTTTTGAAGGCCTATCCTCTGCTCTGTATCTCAATTACTGTTCTTCA 148

Qy    421 tttcaattattcttacctactattcagttcccttgatcttttcttcttgggggctgtctt 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 TTTCAATTATTCTTACCTACTATTCAGTTCCCTTGATCTTTTCTTCTTGGGGGCTGTCTT 88

Qy    481 agggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 AGGGTCAGGGAGATTGCAGAAGCACCAGAACTAGGAGCAGCCCTGAGACATGGGGAGTTG 28

Qy    541 gagctgaaggaggaatggcaggatgaa 567
          |||||||||||||||||||||||||||
```

```
Db      27 GAGCTGAAGGAGGAATGGCAGGATGAA 1


RESULT   4
BF330839/c
LOCUS       BF330839      369 bp    mRNA            EST       22-NOV-2000
DEFINITION  RC3-BT0333-110900-116-a06 BT0333 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF330839
VERSION     BF330839.1  GI:11301587
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC3&t2=RC3-BT0333-
            110900-116-a06&t3=2000-09-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 369.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0333"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA,and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      117 a     78 c     92 g     82 t
ORIGIN


  Query Match           46.4%;  Score 296.4;  DB 146;  Length 369;
  Best Local Similarity   97.8%;  Pred. No. 4.8e-74;
  Matches  311;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy     180 ctgggttccagccacttgtaatgtgggacatctctcaccccaactttgtaggtggagcaa 239
           || |   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 CTTGCACCCAGCCACTTGTAATGTGGGACATCTCTCACCCCAACTTTGTAGGTGGAGCAA 273

Qy     240 ctgctacagaggtaaatatgattaactttacattccatctttcgtctgctcccaaacttta 299
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     272 CTGCTACAGAGGTAAATATGATTAACTTTATATTCCATCTTTCGTCTGCTCCCAAACTTA 213

Qy     300 acagcaggtaatctgcttctagcaagtggtgaaggtaagagaagcatctgtataggaggc 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 ACAGCAGGTAATCTGCTTCTAGCAAGTGGTGAAGGTAAGAGAAGCATCTGTATAGGAGGC 153

Qy     360 aagagatctgagtccttttgaaggcctatcctctgctctgtatctcaattactgttcttc 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152 AAGAGATCTGAGTCCTTTTGAAGGCCTATCCTCTGCTCTGTATCTCAATTACTGTTCTTC 93

Qy     420 atttcaattattcttacctactattcagttcccttgatcttttcttcttgggggctgtct 479
           |||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      92 ATTTGAATTATTCTTACCTACTATTC-GTTCCCTTGATCTTTTCTTCTTGGGGGCTGTCT 34

Qy     480 tagggtcagggagattgc 497
           ||||||||||||||||||
Db      33 TAGGGTCAGGGAGATTGC 16


RESULT   5
AW806456
LOCUS       AW806456      229 bp    mRNA            EST       16-MAY-2000
DEFINITION  MR4-UM0115-040500-203-d06 UM0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW806456
VERSION     AW806456.1  GI:7858326
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR4-UM0115-040
            500-203-d06&t3=2000-05-04&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 43
            High quality sequence stop: 229.
FEATURES             Location/Qualifiers
     source          1. .229
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UM0115"
                     /dev_stage="Adult"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       76 a     43 c     73 g     37 t
ORIGIN


  Query Match           15.3%;  Score 97.8;  DB 120;  Length 229;
  Best Local Similarity   89.7%;  Pred. No. 2.6e-17;
  Matches  105;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     477 tcttagggtcagggagattgcagaagcaccagaactaggagcagccctgagacatgggga 536
           | || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    103 TGTTGGGGTCAGGGAGATTGCAGAAGCACCAGAACTAGGAGCAGCCCTGAGACATGGGGA 162

Qy    537 gttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgag 593
          ||||||||||||||||||||||||||||||||||||||  |  ||  |  |||
Db    163 GTTGGAGCTGAAGGAGGAATGGCAGGATGAAGAATTCCCTTAGATAGCTCCCTAGAG 219


RESULT   6
AW806459/c
LOCUS       AW806459      148 bp    mRNA            EST       16-MAY-2000
DEFINITION  MR4-UM0115-040500-203-h07 UM0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW806459
VERSION     AW806459.1  GI:7858329
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR4-UM0115-040
            500-203-h07&t3=2000-05-04&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 36
            High quality sequence stop: 142.
FEATURES             Location/Qualifiers
     source          1. .148
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UM0115"
                     /dev_stage="Adult"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       25 a     48 c     27 g     48 t
ORIGIN


  Query Match             9.8%;  Score 62.8;  DB 120;  Length 148;
  Best Local Similarity   86.2%;  Pred. No. 2.4e-07;
  Matches   81;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy    488 gggagattgcagaagcaccagaactaggagcagccctgagacatggggagttggagctga 547
          ||   | |||||  ||||||||||||  ||||||||||||||||||  |||||||  ||||||||
Db    101 GGAGAACTGCAGTAGCACCAGAACATGGAGCAGCCCTGAGACA-GGGGAGTCGGAGCTGA 43

Qy    548 aggaggaatggcaggatgaagaattccctaggtg 581
          ||||||||||||||||||||  || | ||||  ||
Db     42 AGGAGGAATGGCAGGATGAAGTATCCACTAGATG 9


RESULT   7
BE284691
LOCUS       BE284691      473 bp    mRNA            EST       26-OCT-2000
DEFINITION  601098104F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3496753 5',
            mRNA sequence.
ACCESSION   BE284691
VERSION     BE284691.1  GI:9161804
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8549  row: h  column: 02
            High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3496753"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      123 a    122 c    150 g     78 t
ORIGIN


  Query Match             9.7%;  Score 62.2;  DB 165;  Length 473;
  Best Local Similarity   76.8%;  Pred. No. 5e-07;
  Matches   76;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy    482 gggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttgg 541
          |||  | |||||  ||||  |||  |||||| |||||||  ||  |||  ||  |||  |||  |||
Db    123 GGGCTAAGGAGAAGGCAGGAGCCTCAGAACCAGGAGCAACCATGAAACTTGGAGAGCTGG 182

Qy    542 agctgaaggaggaatggcaggatgaagaattccctaggt 580
          ||||  |  ||||| |||||  |||||||| ||  ||||| |
Db    183 AGCTAAGAGAGGAGTGGCAAGATGAAGAGTTTCCTAGAT 221


RESULT   8
BF021610
LOCUS       BF021610      519 bp    mRNA            EST       29-DEC-2000
DEFINITION  uy49f05.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3662913 5'
            similar to SW:NIP2_MOUSE O54940 BCL2/ADENOVIRUS E1B 19-KDA
            PROTEIN-INTERACTING PROTEIN 2. [1] ;, mRNA sequence.
ACCESSION   BF021610
VERSION     BF021610.1  GI:10752942
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1423681
            Seq primer: -40RP from Gibco
            High quality sequence stop: 427.
FEATURES              Location/Qualifiers
     source           1. .519
                      /organism="Mus musculus"
                      /strain="CZECH II"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3662913"
                      /clone_lib="NCI_CGAP_Lu30"
                      /tissue_type="tumor, metastatic to mammary"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; transgenic model WNT-1, expression driven by
                      MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                      dT.   Library constructed by Life Technologies.
                      Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      136 a    136 c    163 g     84 t
ORIGIN


  Query Match             9.5%;  Score 60.6;  DB 142;  Length 519;
  Best Local Similarity   75.8%;  Pred. No. 1.5e-06;
  Matches   75;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     482 gggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttgg 541
           |||    |||||  |||| |||  |||||| ||||||| || ||| || ||| ||| |||
Db     158 GGGCTGAGGAGAAGGCAGGAGCCTCAGAACCAGGAGCAACCATGAAACTTGGAGAGCTGG 217

Qy     542 agctgaaggaggaatggcaggatgaagaattccctaggt 580
           |||| |  ||||| ||||| |||||||| || ||||| |
Db     218 AGCTAAGAGAGGAGTGGCAAGATGAAGAGTTTCCTAGAT 256


RESULT  9
BE062071/c
LOCUS       BE062071      411 bp    mRNA            EST       09-JUN-2000
DEFINITION  RC1-BT0254-160300-018-d03 BT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE062071
VERSION     BE062071.1  GI:8406721
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC1-BT0254-160
            300-018-d03&t3=2000-03-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 411.
FEATURES              Location/Qualifiers
     source           1. .411
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0254"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT      119 a     94 c     90 g    108 t
ORIGIN


  Query Match             8.7%;  Score 55.8;  DB 162;  Length 411;
  Best Local Similarity   95.8%;  Pred. No. 3.2e-05;
  Matches   68;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 ccagaaccgagtttaggtccaggttctcgttctggcaaatctttctccttaccttcttcc 60
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      94 CCAGAACCGAGTTTAGGTCCAGGTTCTCGTTCTGGC-AATCTTTCTCCTTACCTTCTTCC 36

Qy      61 tccacccctcc 71
           |||||||  ||
Db      35 TCCACCCTGCC 25


RESULT  10
BG066225
LOCUS       BG066225     1451 bp    mRNA            EST       26-JAN-2001
DEFINITION  H3040D10-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3040D10 3', mRNA sequence.
ACCESSION   BG066225
VERSION     BG066225.1  GI:12548794
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1451)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3040  row: D  column: 10
            Seq primer: -21M13 Forward
            High quality sequence stop: 1451
            POLYA=Yes.
```

```
FEATURES             Location/Qualifiers
     source          1. .1451
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="H3040D10"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers. References include: (1) Genome-wide
                     expression profiling of mid-gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA analysis reveals phased gene expression
                     patterns during preimplantation mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day mouse embryos reveals enrichment in the t-complex
                     and under-representation on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT       283 a    258 c    694 g    216 t
ORIGIN
```

```
  Query Match             6.9%;  Score 44.2;  DB 172;  Length 1451;
  Best Local Similarity   53.8%;  Pred. No. 0.097;
  Matches   91;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           ||| | || |  |||| | |||||   |  |        | |   || || |    | ||
Db    1190 GGGTGGTGCCGGAGGGGCGGGGAGGGGGGGGGAGGAGGGGAGGGGGGGCGGGGAGGGGA 1249

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
           |  |||  | ||| || |  || ||   ||  ||| |  |    |  ||  |||| |
Db    1250 CGCGGGTGGGTGGGGCGGGGGGGGGCCGGGGGGAGGGGGGGGGCGGGAGAGGAGGGGGC 1309

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
           | | ||| || |||| |  ||| |||||||   |||  ||   | ||||
Db    1310 GGGGGGGGGGGTGGGGGGGGGGGGGGGTGGGGGCGAGGGGGCGCGGGG 1358
```

```
RESULT  11
CNS016AN
LOCUS       CNS016AN     1201 bp     DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106505
VERSION     AL106505.1  GI:5622253
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN15E02"
                     /note="end : T7"
BASE COUNT       227 a    109 c    261 g    439 t    165 others
ORIGIN
```

```
  Query Match             6.8%;  Score 43.4;  DB 229;  Length 1201;
  Best Local Similarity   34.4%;  Pred. No. 0.16;
  Matches  105;  Conservative   49;  Mismatches  151;  Indels    0;  Gaps    0;

Qy     329 tgaaggtaagagaagcatctgtataggaggcaagagatctgagtccttttgaaggcctat 388
           |: |||   :  :||  | |  :   :  : ::|||   || |  ||:::|   : : |
Db     834 TRGAGGAGGKTAMAGGGTGTTAWSTTAMATYGMRAGAAKGGAWTRTTTWWKATTTMRKTT 893

Qy     389 cctctgctctgtatctcaattactgttcttcatttcaattattcttacctactattcagt 448
             |:| :: |   | : | ||| ::|: || |||:   || || ||   |  | ||   |
Db     894 AWTMTAHWTTAATTAWTATTTATWKTKATTTATTWTTTTTTTTTTTTTTTTTTTTTTTT 953

Qy     449 tcccttgatcttttcttcttgggggctgtcttagggtcagggagattgcagaagcaccag 508
           |   ||  | ||:| |: || :|:: ::| || :  |         | |  |: :     |
Db     954 TATTTTTTTTTTKTGTKTTTTKGRKTKKTTTTTKTTTTGTTTTTKTGGGGGRGKGGGGGG 1013

Qy     509 aactaggagcagccctgagacatggggagttggagctgaaggaggaatggcaggatgaag 568
             | |          : |    ||||||  :| |  |  || || : ||   :: || |
Db    1014 TTTTTGTTTTTTTGGGSGGGGGGGGGGAGAGRGTGGGGGGGGGGGGRGGGRGARRAGAGG 1073

Qy     569 aattccctaggtgaggacgtgtgagggtggctgggagaagggaggggtggtcacgaatgg 628
           :       || |:|:    | ||:   |:  | ||:||:|| :::: ||  |  :| |
Db    1074 RTGGGGGGGGGRGRGRGGAAGAGARAAAGRARGAGARAARGGGRRRRGGGAGAGARARGA 1133

Qy     629 acgga 633
           :  ||
Db    1134 RGAGA 1138
```

```
RESULT  12
B21587/c
LOCUS       B21587         869 bp    DNA             GSS       16-SEP-1997
DEFINITION  T29I6-Sp6 TAMU Arabidopsis thaliana genomic clone T29I6, DNA
            sequence.
ACCESSION   B21587
VERSION     B21587.1  GI:2396641
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T29I6-T7
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
```

```
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
            High quality sequence start: 452
            High quality sequence stop: 463.
FEATURES              Location/Qualifiers
     source           1. .869
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="T29I6"
                      /clone_lib="TAMU"
                      /sex="hermaphrodite"
                      /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                      ; Produced by Rod Wing"
BASE COUNT       53 a    439 c    104 g    225 t     48 others
ORIGIN


  Query Match             6.5%;  Score 41.6;  DB 227;  Length 869;
  Best Local Similarity   52.9%;  Pred. No. 0.46;
  Matches   83;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     481 agggtcagggagattgcagaagcaccagaactaggagcagccctgagacatggggagttg 540
           ||||  | |||||  | | ||| |  || |  || |  ||    |||| | ||||||
Db     780 AGGGAGANGGAGAGAGGAAAAGGAGAAGGAGGAGAAAGAGAGGAGAGAGACGGGGAGCGA 721

Qy     541 gagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgtgtgagggtggct 600
           |||  |||||  ||  |  |||  | |  |   |   || |   | | | | ||| |
Db     720 GAGAGGAAGGNAGAGGGCGAGGGGGGAAGAGAGCGAGGGAGGACAGGGGGGCGGGCGCGG 661

Qy     601 gggagaagggaggggtggtcacgaatggacggagggg 637
           || |  ||||| || || |      ||  |  || |
Db     660 CGGGGGNGGGACGGGGGGCCNAGCGGGCNGCGGGCG 624


RESULT  13
AW926672
LOCUS       AW926672      872 bp    mRNA            EST       25-JUL-2000
DEFINITION  HVSMEg0007N11 Hordeum vulgare pre-anthesis spike EST library
            HVcDNA0008 (white to yellow anther) Hordeum vulgare cDNA clone
            HVSMEg0007N11, mRNA sequence.
ACCESSION   AW926672
VERSION     AW926672.1  GI:8098414
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Hordeum.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            High quality sequence stop: 872.
FEATURES              Location/Qualifiers
     source           1. .872
                      /organism="Hordeum vulgare"
                      /cultivar="Morex"
                      /db_xref="taxon:4513"
                      /clone="HVSMEg0007N11"
                      /clone_lib="Hordeum vulgare pre-anthesis spike EST library
                      HVcDNA0008 (white to yellow anther)"
                      /tissue_type="pre-anthesis spike"
                      /lab_host="SOLR"
                      /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1"
BASE COUNT      157 a     99 c    554 g     62 t
ORIGIN


  Query Match             6.5%;  Score 41.6;  DB 121;  Length 872;
  Best Local Similarity   53.0%;  Pred. No. 0.46;
  Matches   89;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           |||||  |     | |  |||||||  |    ||     |||    | || | |  | ||
Db     611 GGGGGGGGGGGGGGAGGGAGGGAGAGGGGCGAGAGGGAGGAAGGGAGCGCGGGCGAGGGA 670

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
           ||||  |  ||||  |||||||||   ||  ||| |  ||       ||| | ||  |
Db     671 GGAGGGGGGGAGGAGAGGAAGGAGGGGGGGGGGGAGGGTGAGAGGGAGAGGGGGGGGGA 730

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggaggg 636
           | ||||| ||  ||| |  ||| |||| ||    |   ||| |  | |
Db     731 GAGAGGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGAGGAGGAG 778


RESULT  14
BF256465
LOCUS       BF256465      1097 bp    mRNA            EST       16-NOV-2000
DEFINITION  HVSMEf0010A21f Hordeum vulgare seedling root EST library HVcDNA0007
            (etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMEf0010A21f, mRNA sequence.
ACCESSION   BF256465
VERSION     BF256465.1  GI:11185578
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Hordeum.
REFERENCE   1  (bases 1 to 1097)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 8
            High quality sequence stop: 1056.
FEATURES              Location/Qualifiers
     source           1. .1097
                      /organism="Hordeum vulgare"
                      /cultivar="Morex"
                      /db_xref="taxon:4513"
                      /clone="HVSMEf0010A21f"
                      /clone_lib="Hordeum vulgare seedling root EST library
                      HVcDNA0007 (etiolated and unstressed)"
                      /tissue_type="Seedling root"
                      /lab_host="TJC121"
                      /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1"
BASE COUNT      169 a     10 c    859 g     45 t     14 others
ORIGIN


  Query Match             6.5%;  Score 41.6;  DB 145;  Length 1097;
```

```
  Best Local Similarity   52.7%;  Pred. No. 0.49;
  Matches   89;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     469 gggggctgtcttagggtcagggagattgcagaagcaccagaactaggagcagccctgaga 528
           |||||  |   | |||   ||| |   |  |  |     |   | || |  |    | |
Db     200 GGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGG 259

Qy     529 catggggagttggagctgaaggaggaatggcaggatgaagaattccctaggtgaggacgt 588
             ||||    || |  |  || |||| || ||    |||||||   | || | ||| |
Db     260 GGGGGGGGTGGGGGGAGGGGGGGGGAAAGGAAGAGAAAAGAATTTTGTTGGGGGGGAGGG 319

Qy     589 gtgagggtggctgggagaagggaggggtggtcacgaatggacggagggg 637
           | | ||| ||  ||| |  ||| |||| ||    ||| ||  || ||||
Db     320 GGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGAAGGGGGGGGGGGG 368


RESULT  15
AL513839
LOCUS       AL513839      307 bp    mRNA            EST       13-FEB-2001
DEFINITION  AL513839 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA002ZA09 3
            prime, mRNA sequence.
ACCESSION   AL513839
VERSION     AL513839.1  GI:12777333
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .307
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CL0BA002ZA09"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT        0 a      5 c    164 g    114 t     24 others
ORIGIN


  Query Match             6.4%;  Score 41.2;  DB 105;  Length 307;
  Best Local Similarity   42.1%;  Pred. No. 0.44;
  Matches  106;  Conservative   19;  Mismatches  127;  Indels    0;  Gaps    0;

Qy     386 tatcctctgctctgtatctcaattactgttcttcatttcaattattcttacctactattc 445
           | | | | | | | | |   ||  | || ||  |||   || || ||   |  | ||
Db      36 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 95

Qy     446 agttcccttgatctttttcttcttgggggctgtcttagggtcagggagattgcagaagcac 505
             ||   ||  | |||| || |::: :: ::: |: :::: ::: : ::|  |  |
Db      96 TTTTTTTTTTTTTTTTTTTTTTKKKTKKKKKKKTTKTKKKKKKKKKKKKKKKGGGGGGGGGG 155

Qy     506 cagaactaggagcagccctgagacatggggagttggagctgaaggaggaatggcaggatg 565
             |     || |  |    | |    |||| |  || |  |  || ||   ||  ||  |
Db     156 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 215

Qy     566 aagaattccctaggtgaggacgtgtgagggtggctgggagaagggaggggtggtcacgaa 625
           |          || | ||  | | | ||| ||  ||| |  ||| |||| ||      |
Db     216 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 275

Qy     626 tggacggagggg 637
             ||  || ||||
Db     276 GGGGGGGGGGGG 287


Search completed: April 30, 2001, 12:17:47
Job time: 12536 sec
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:14:51 ; Search time 22667.1 Seconds
(without alignments)
211.629 Million cell updates/sec

Title: US-09-721-183-5
Perfect score: 779
Sequence: 1 gtatacattctttattaatc.....gaaactgtcagaagcaaaaa 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 3078962356 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	622	79.8	1030	9	AX053806	AX053806 Sequence
2	583.2	74.9	992	9	AX053824	AX053824 Sequence
3	557.4	71.6	904	9	AX053778	AX053778 Sequence
4	530.2	68.1	664	9	AX053793	AX053793 Sequence
5	512.2	65.8	814	9	AX053780	AX053780 Sequence
6	507.8	65.2	820	9	AX053783	AX053783 Sequence
7	503.8	64.7	1158	9	AX053798	AX053798 Sequence
8	487.8	62.6	713	9	AX053787	AX053787 Sequence
9	479.8	61.6	776	9	AX053782	AX053782 Sequence
10	464.6	59.6	741	9	AX053804	AX053804 Sequence
11	456.8	58.6	708	9	AX053785	AX053785 Sequence

	12	425.2	54.6	833	9	AX053800	AX053800 Sequence
	13	417.8	53.6	636	9	AX053794	AX053794 Sequence
	14	380	48.8	776	9	AX053808	AX053808 Sequence
	15	334.8	43.0	771	9	AX053802	AX053802 Sequence
	16	318.4	40.9	766	9	AX053799	AX053799 Sequence
	17	236	30.3	609	9	AX053810	AX053810 Sequence
c	18	196.6	25.2	756	9	AX053803	AX053803 Sequence
c	19	154	19.8	175917	80	AL357148	AL357148 Homo sapi
c	20	150.8	19.4	185869	81	AL391494	AL391494 Homo sapi
	21	131.6	16.9	111648	91	HS16C2	AL049911 Homo sapi
	22	131.6	16.9	119182	92	HSB62L20	AL050302 Homo sapi
c	23	130	16.7	340000	91	HS21C003	AL163203 Homo sapi
	24	126.8	16.3	144096	80	AL359312	AL359312 Homo sapi
	25	124.6	16.0	159455	81	AL512324	AL512324 Homo sapi
	26	124.6	16.0	208898	80	AL358394	AL358394 Homo sapi
	27	116.8	15.0	175917	80	AL357148	AL357148 Homo sapi
	28	114.2	14.7	157048	88	AC067744	AC067744 Homo sapi
	29	114.2	14.7	185869	81	AL391494	AL391494 Homo sapi
c	30	111.2	14.3	148267	81	AL390758	AL390758 Homo sapi
	31	108.8	14.0	1805	88	AK001137	AK001137 Homo sapi
	32	108.8	14.0	5360	85	AB028997	AB028997 Homo sapi
c	33	95.2	12.2	177132	60	AC008386	AC008386 Homo sapi
c	34	93.2	12.0	154904	89	AL162272	AL162272 Human DNA
c	35	93.2	12.0	194832	81	AL365501	AL365501 Homo sapi
c	36	88.4	11.3	162651	80	AL356259	AL356259 Homo sapi
c	37	88.4	11.3	164118	67	AC021384	AC021384 Homo sapi
c	38	88.4	11.3	183684	79	AL162715	AL162715 Homo sapi
c	39	83	10.7	169054	70	AC026271	AC026271 Homo sapi
	40	83	10.7	169198	71	AC027266	AC027266 Homo sapi
c	41	83	10.7	192337	87	AC022596	AC022596 Homo sapi
	42	81.8	10.5	1146	10	I40055	I40055 Sequence 1
	43	81.8	10.5	1146	93	HSU28831	U28831 Human prote
	44	81.8	10.5	2924	85	AB046861	AB046861 Homo sapi
c	45	80.8	10.4	123214	76	AC079144	AC079144 Homo sapi

ALIGNMENTS

```
RESULT   1
AX053806
LOCUS       AX053806      1030 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 572 from Patent WO0073801.
ACCESSION   AX053806
VERSION     AX053806.1  GI:12228125
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1030)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 572 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .1030
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      383 a    170 c    219 g    258 t
ORIGIN
```

Query Match 79.8%; Score 622; DB 9; Length 1030;
Best Local Similarity 98.4%; Pred. No. 5.6e-120;
Matches 628; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy    141 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCAC 60

Qy    201 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAAGGATGTGTGTTTACCCAAGGCTACACATCAAAAAGAAATAGATAAAATAAATGGAA 120

Qy    261 aattagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagttt 320
          |||||||||  ||||||||  |||  ||||||  ||||||||||   ||||  ||||||||||
Db    121 AATTAGAAGAGTCTCCTGATAATGATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTT 180

Qy    321 ctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccg 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCG 240

Qy    381 agaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagcct 440
          ||||||||||||||  |  ||||||||||||||||||||||||||||||||||||||||||
Db    241 AGAAGCCATCTGCCATTTAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCT 300

Qy    441 tggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaac 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGAATTGAAGAATGAACAAACATTGAGAGCAGATCAGATGTTCCCTTCAGAATCAAAAC 360

Qy    501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAAAGAAGGTTGAAGAAAATTCTTGGGATTCTGAGAGTCTCCGTGAGACTGTTTCACAGA 420

Qy    561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTGTGTGTACCCAAGGCTACACATCAAAAAGAAATGGATAAAATAAGTGGAAAAT 480

Qy    621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TAGAAGATTCAACTAGCCTATCAAAAATCTTGGATACAGTTCATTCTTGTGAAAGAGCAA 540

Qy    681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGGAACTTCAAAAAGATCACTGTGAACAACGTACAGGAAAAATGGAACAAATGAAAAAGA 600

Qy    741 agttttgtgtactgaaaaagaaactgtcagaagcaaaa 778
          |||||||||||||||||||||||||||||||||||||
Db    601 AGTTTTGTGTACTGAAAAAGAAACTGTCAGAAGCAAAA 638
```

```
RESULT   2
AX053824
LOCUS       AX053824      992 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 590 from Patent WO0073801.
ACCESSION   AX053824
VERSION     AX053824.1  GI:12228143
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 590 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      348 a    163 c    200 g    243 t     38 others
ORIGIN
```

Query Match 74.9%; Score 583.2; DB 9; Length 992;
Best Local Similarity 94.7%; Pred. No. 6.7e-112;
Matches 605; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

```
Qy    141 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCAC 60
```

```
Qy    201 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAAGGATGTGTGTTTACCCAAGGCTACACATCAAAAAGAAATAGATAAAATAAATGGAA 120

Qy    261 aattagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagttt 320
          |||||  |||  ||||||||  |||  ||||||  ||||||||||  ||||  ||||||||||
Db    121 AATTANAAGAGTCTCCTGATAATGATGGTTTTCTGAAGGCTCCCTGCANAATGAAAGTTT 180

Qy    321 ctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccg 380
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db    181 CTATTCCAACTAAAGCCTTANAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCG 240

Qy    381 agaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagcct 440
          |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGAAGCCNTCTGCCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCT 300

Qy    441 tggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaac 500
          ||||||||||||||||  ||||||||||||||||||||||||||||||  |||||||||
Db    301 TGGAATTGAAGAATGAACNAACATTGAGAGCAGATCAGATGTTCCCTTCANAATCAAAAC 360

Qy    501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
          |||||||||  |||||||||||||||||||||||  |||||||  ||||||||||||||||
Db    361 AAAAGAAGGNTGAAGAAAATTCTTGGGATTCTGANAGTCTCCNTGAGACTGTTTCACAGA 420

Qy    561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGGATGTGTGTGTACCCAAGGCTACACATCAAAAAGAAATGGATAAAATAAGTGGAAAAT 480

Qy    621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db    481 TAGAAGATTCAACTAGCCTATCAAAAATCTTGGATACAGTTCATTCTTGTGAAAGANCAA 540

Qy    681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
          ||||||||  ||||||||      |  | |||||  ||||||||||||||||||||||||
Db    541 NGGAACTTCNAAAAGATCCTGTGAACAACCGTACNNGAAAAATGGAACAAATGAAAAAGA 600

Qy    741 agtttt-gtgtactgaaaaagaaactgtcagaagcaaaa 778
          ||||||  ||||  ||||||||||||||||  |||||  |||
Db    601 AGTTTTGGTGTNCTGAAAAAGAAACTGTCCGAAGCCAAA 639


RESULT   3
AX053778
LOCUS       AX053778      904 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 544 from Patent WO0073801.
ACCESSION   AX053778
VERSION     AX053778.1  GI:12228097
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 904)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 544 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      315 a    173 c    200 g    216 t
ORIGIN


  Query Match             71.6%;  Score 557.4;  DB 9;  Length 904;
  Best Local Similarity   95.7%;  Pred. No. 1.6e-106;
  Matches  573;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          ||  |  |||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db     46 CCTTCGAGCCTGCCATTGATATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 105

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    106 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 165

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166 AAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 225

Qy    216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
          ||||||||||  | ||||||||||||||||||||||||||||||||||||||||||||
Db    226 TACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGGTCTC 285

Qy    276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 CTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAG 345

Qy    336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 CCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCT 405

Qy    396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 TCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATG 465

Qy    456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
          ||||||||||||||||||  ||||  ||||  ||||||||  ||||||||  |   |||||
Db    466 AACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAG 525

Qy    516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
          |||  ||||||||||||||||||||  ||||||||||||||||||||||||||||||  |||
Db    526 AAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTAC 585

Qy    576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaact 634
          ||||||||  |  ||||||||||||||  |||||||||  ||||||||||||||  ||  ||
Db    586 CCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGGTCTCCT 644


RESULT   4
AX053793
LOCUS       AX053793      664 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 559 from Patent WO0073801.
ACCESSION   AX053793
VERSION     AX053793.1  GI:12228112
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 559 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      236 a    115 c    144 g    166 t      3 others
ORIGIN


  Query Match             68.1%;  Score 530.2;  DB 9;  Length 664;
  Best Local Similarity   95.3%;  Pred. No. 7.2e-101;
  Matches  566;  Conservative    0;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          ||  |  |||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db      4 CCTTCGAGCCTGCCATTGATATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 63

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      64 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 123

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 AAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 183

Qy     216 tacccaaggctacacatcaaaaagaaatagatataaaataaatggaaaattagaagggtctc 275
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db     184 TACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGGTCTC 243

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 CTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAG 303

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 CCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCT 363

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 TCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATG 423

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           ||||||||||||||||||| |||| |||| |||||||| ||||||||    |||||
Db     424 AACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAG 483

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||| |||||||||||||||||||| |||||||||||||| ||||||||||||||||| |||
Db     484 AAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTC-CAGAAGGATGTGTGTTTAC 542

Qy     576 ccaaggc-tacacatcaaaaagaaatggataaaataagtggaaaattagaagat 628
           |||| || | | | |||||||||||| |||||||||| ||||||||||||  |
Db     543 CCAANGCTTGCGCNTCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAANGT 596


RESULT   5
AX053780
LOCUS       AX053780      814 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 546 from Patent WO0073801.
ACCESSION   AX053780
VERSION     AX053780.1  GI:12228099
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 546 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      278 a    159 c    173 g    195 t      9 others
ORIGIN


  Query Match           65.8%;  Score 512.2;  DB 9;  Length 814;
  Best Local Similarity   93.7%;  Pred. No. 4e-97;
  Matches  554;  Conservative    0;  Mismatches   35;  Indels    2;  Gaps    2;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 CCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 90

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 150

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 AAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 210

Qy     216 tacccaaggctacacatcaaaaagaaatagatataaaataaatggaaaattagaagggtctc 275
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db     211 TACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGGTCTC 270

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     271 CTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAG 330

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     331 CCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCCATCTGCT 390

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391 TCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATG 450

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           |||||||||||||||||||| |||| |||| |||||||| ||||||||    |||||
Db     451 AACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAG 510

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||| |||||||||||||||||||| |||||||||||||| || |||||||||||| | |
Db     511 AAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTC-CANAAGGATGTGTGT-TTC 568

Qy     576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaag 626
           |||||||| | |||||||||||||| ||| |||||| |||||  |    | |
Db     569 CCAAGGCTGCGCATCAAAAAGAAATAGATTAAATAATGGGAAATTAGAANG 619


RESULT   6
AX053783
LOCUS       AX053783      820 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 549 from Patent WO0073801.
ACCESSION   AX053783
VERSION     AX053783.1  GI:12228102
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 549 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      282 a    155 c    168 g    203 t     12 others
ORIGIN


  Query Match           65.2%;  Score 507.8;  DB 9;  Length 820;
  Best Local Similarity   98.5%;  Pred. No. 3.3e-96;
  Matches  523;  Conservative    0;  Mismatches    7;  Indels    1;  Gaps    1;

Qy      94 gaatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggacta 153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTA 60

Qy     154 tgaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtg 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTG 120

Qy     214 tttacccaaggctacacatcaaaaagaaatagatataaaataaatggaaaattagaagggtc 273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     121 TTTACCCAAGGCTACACATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGAGTC 180
```

```
Qy    274 tcctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaa 333
          ||||| ||| |||||| |||||||||||  |||| ||||||||||||||||||||||||||
Db    181 TCCTGATAATGATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTTCTATTCCAACTAA 240

Qy    334 agccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgc 393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGC 300

Qy    394 cttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaa 453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAA 360

Qy    454 tgaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttga 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGAACAAACATTGAGAGCAGATCAGATGTTCCCTTCAGAATCAAAACAAAAGAAGGTTGA 420

Qy    514 agaaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgt 573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGAAAATTCTTGGGATTCTGAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTGT 480

Qy    574 acccaaggctacacatcaaaaagaaatggataaaataagtggaaaattaga 624
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 -CCCAAGGCTACACATCAAAAAGAAATGGATAAAATAAGTGGAAAATTAGA 530


RESULT  7
AX053798
LOCUS       AX053798      1158 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 564 from Patent WO0073801.
ACCESSION   AX053798
VERSION     AX053798.1  GI:12228117
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1158)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 564 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .1158
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      399 a    216 c    241 g    300 t      2 others
ORIGIN


  Query Match           64.7%;  Score 503.8;  DB 9;  Length 1158;
  Best Local Similarity   89.9%;  Pred. No. 2.3e-95;
  Matches  569;  Conservative    0;  Mismatches   22;  Indels   42;  Gaps    1;

Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 301

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 361

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 421

Qy    216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaag------ 269
          |||||||||||| | ||||||||||||||||||||||||||||||||||||||      |
Db    422 TACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGTAGAT 481

Qy    270 ----------------------------------ggtctcctgttaaagatggtcttc 293
                                              |||||||||||||||||||||||||
Db    482 ATGCTGCTGAATTTAGAACATTCTCTGCAATGATAAGGTCTCCTGTTAAAGATGGTCTTC 541

Qy    294 tgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggaca 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 TGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGATGGACA 601

Qy    354 tgcaaactttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaa 413
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    602 TGCAAACTTTCAAAGCAGAACCTCCCGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAA 661

Qy    414 tgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcag 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 TGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACATTGAGAGCAG 721

Qy    474 atcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctg 533
          || ||||  |||| |||||||| |||||||| |  |||||||| ||||||||||||||||
Db    722 ATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTG 781

Qy    534 agagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaa 593
          |||||||| |||||||||||||||||||||||||||||| |||||||||| | |||||||
Db    782 AGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAA 841

Qy    594 aagaaatggataaaataagtggaaaattagaag 626
          ||||||| |||||||||| ||||||||||||||
Db    842 AAGAAATAGATAAAATAAATGGAAAATTAGAAG 874


RESULT  8
AX053787
LOCUS       AX053787      713 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 553 from Patent WO0073801.
ACCESSION   AX053787
VERSION     AX053787.1  GI:12228106
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 553 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .713
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      250 a    133 c    153 g    172 t      5 others
ORIGIN


  Query Match           62.6%;  Score 487.8;  DB 9;  Length 713;
  Best Local Similarity   95.5%;  Pred. No. 4.8e-92;
  Matches  512;  Conservative    0;  Mismatches   23;  Indels    1;  Gaps    1;

Qy     95 aatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactat 154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTAT 60

Qy    155 gaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgt 214
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGT 120

Qy    215 ttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtct 274
          |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGGTCT 180

Qy    275 cctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaa 334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAA 240

Qy    335 gccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcc 394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 GCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCC 300

Qy    395 ttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaat 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAAT 360

Qy    455 gaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaa 514
          |||||||||||||||||||| ||||  |||| |||||||| ||||||||| |   ||||
Db    361 GAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAA 420

Qy    515 gaaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgta 574
          ||||  |||||||||||||||||||||| |||||||||||||| |||||||||||| |
Db    421 GAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACANAAGGATGTGTGT-TT 479

Qy    575 cccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattc 630
          ||||||||| | |||||||||||||| |||||||||| |||||||||| | |  ||
Db    480 CCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAAAGGGCTC 535


RESULT  9
AX053782
LOCUS       AX053782      776 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 548 from Patent WO0073801.
ACCESSION   AX053782
VERSION     AX053782.1  GI:12228101
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 548 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      273 a    142 c    167 g    188 t      6 others
ORIGIN


  Query Match           61.6%;  Score 479.8;  DB 9;  Length 776;
  Best Local Similarity   95.7%;  Pred. No. 2.2e-90;
  Matches  493;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    120 agatactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgaga 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGA 60

Qy    180 gtctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctacacatcaaaaag 239
          |||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db     61 GTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAG 120

Qy    240 aaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctgaagg 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAATAGATAAAATAAATGGAAAATTAGAAGGGTCTCCTGTTAAAGATGGTCTTCTGAAGG 180

Qy    300 ctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggacatgcaaa 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAA 240

Qy    360 ctttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaatgcaaa 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAA 300

Qy    420 agtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcagatcaga 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    301 AGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACATTGAGAGCAGATGAGA 360

Qy    480 tgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctgagagtc 539
          |  |||| |||||||| ||||||||| |   |||||||| ||||||||||||||||||||
Db    361 TACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTC 420

Qy    540 tccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaaaagaaa 599
          || ||||||||||||||||||||||||||||| |||||||||| | ||||||||||||
Db    421 TCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGAAA 480

Qy    600 tggataaaataagtggaaaattagaagattcaact 634
          | |||||||||| ||||||||||||||  ||  ||
Db    481 TAGATAAAATAAATGGAAAATTAGAAGGGTCTCCT 515


RESULT  10
AX053804
LOCUS       AX053804      741 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 570 from Patent WO0073801.
ACCESSION   AX053804
VERSION     AX053804.1  GI:12228123
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 570 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      256 a    129 c    160 g    187 t      9 others
ORIGIN


  Query Match           59.6%;  Score 464.6;  DB 9;  Length 741;
  Best Local Similarity   85.7%;  Pred. No. 3.3e-87;
  Matches  565;  Conservative    0;  Mismatches   50;  Indels   44;  Gaps    3;

Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 CCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 71

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 131

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 AAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 191

Qy    216 tacccaaggctacacatcaaaaagaaatagatataaaataaatggaaaattagaag------ 269
          |||||||||||  | ||||||||||||||||||||||||||||||||||||||||||
Db    192 TACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGGTAGAT 251

Qy    270 ------------------------------------ggtctcctgttaaagatggtcttc 293
                                               |||||||||||||||||||||||
Db    252 ATGCTGCTGAATTTGGAACATTCTCTGCAATGATAAGGTCTCCTGTTAAAGATGGTCTTC 311

Qy    294 tgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggaca 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 TGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGATGGACA 371

Qy    354 tgcaaactttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaa 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 TGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAA 431

Qy    414 tgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcag 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 TGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACATTGAGAGCAG 491
```

```
Qy    474 atcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctg 533
          || ||||  |||| |||||||| ||||||||| |   |||| ||| ||||||||||||||
Db    492 ATGAGAT-CTCCCATCAGAATCCAAACAAAAGGACTATGAAAAAAGTTCTTGGGATTCTG 550

Qy    534 agagtctccgtgagactgtttcacag-aaggatgtgtgtgtacccaaggctacacatcaa 592
          |||||||| |||||||||||||||| |||||||||||| ||||| | | | | | ||||
Db    551 AGAGTCTCTGTGAGACTGTTTCACAGAAAGGATGTGTGTTTACCCCAAGNTGCGCNTCAA 610

Qy    593 aaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatctt 651
          || ||||| ||| |||| | ||| ||| | |  |      |   |   |||   |||
Db    611 AAGGAAATTGATTAAATTAATGGGAAAATTTANAANGGTCTTCTGGTNAAAAAGGNCTT 669


RESULT 11
AX053785
LOCUS       AX053785      708 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 551 from Patent WO0073801.
ACCESSION   AX053785
VERSION     AX053785.1  GI:12228104
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 551 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      246 a    134 c    154 g    172 t      2 others
ORIGIN


  Query Match           58.6%;  Score 456.8;  DB 9;  Length 708;
  Best Local Similarity   95.5%;  Pred. No. 1.4e-85;
  Matches  470;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    143 caaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcacag 202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAG 60

Qy    203 aaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaa 262
          |||||||||||||||||||||| | ||||||||||||||||||||||||||||||||||
Db     61 AAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAA 120

Qy    263 ttagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttct 322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTAGAAGGGTCTCCTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCT 180

Qy    323 attccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccgag 382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAG 240

Qy    383 aagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttg 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTG 300

Qy    443 gaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaa 502
          |||||||||||||||||||||||||||||||| ||||  |||| |||||||| ||||||
Db    301 GAATTGAAGAATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAA 360

Qy    503 aagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacagaag 562
          ||| |   |||||||| |||||||||||||||||||| |||||||||||||||||||||
Db    361 AAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAG 420

Qy    563 gatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaatta 622
          ||||||||| |||||||||| | |||||||||||||| |||||||||| ||||||||||
Db    421 GATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAAATTA 480

Qy    623 gaagattcaact 634
          ||||  ||  ||
Db    481 GAAGGGTCTCCT 492


RESULT 12
AX053800
LOCUS       AX053800      833 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 566 from Patent WO0073801.
ACCESSION   AX053800
VERSION     AX053800.1  GI:12228119
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 566 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .833
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      283 a    154 c    182 g    206 t      8 others
ORIGIN


  Query Match           54.6%;  Score 425.2;  DB 9;  Length 833;
  Best Local Similarity   89.1%;  Pred. No. 5.2e-79;
  Matches  488;  Conservative    0;  Mismatches   18;  Indels   42;  Gaps    1;

Qy    121 gatactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgagag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAG 60

Qy    181 tctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctacacatcaaaaaga 240
          |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db     61 TCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGA 120

Qy    241 aatagataaaataaatggaaaattagaag---------------------------- 269
          |||||||||||||||||||||||||||||
Db    121 AATAGATAAAATAAATGGAAAATTAGAAGGTAGATATGCTGCTGAATTTAGAACATTCTC 180

Qy    270 ----------ggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagt 318
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGCAATGATAAGGTCTCCTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGT 240

Qy    319 ttctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcc 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCTATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCC 300

Qy    379 cgagaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagc 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGC 360

Qy    439 cttggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaa 498
          |||||||||||||||||||||||||||||||||||| ||||  |||| |||||||| ||
Db    361 CTTGGAATTGAAGAATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAA 420

Qy    499 acaaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcaca 558
          ||||||| |   |||||||| |||||||||||||||||||| |||||||||||||||||
Db    421 ACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACA 480

Qy    559 gaaggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaa 618
          |||||||||||| |||||||||| | |||||||||||||| |||||||||| ||||||
Db    481 GAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAA 540

Qy    619 attagaag 626
          ||||||||
Db    541 ATTAGAAG 548
```

```
Db    541 ATTAGAAG 548


RESULT 13
AX053794
LOCUS       AX053794      636 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 560 from Patent WO0073801.
ACCESSION   AX053794
VERSION     AX053794.1  GI:12228113
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 560 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      217 a    120 c    138 g    157 t      4 others
ORIGIN


  Query Match             53.6%;  Score 417.8;  DB 9;  Length 636;
  Best Local Similarity   94.7%;  Pred. No. 1.8e-77;
  Matches  431;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy    180 gtctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctacacatcaaaaag 239
          ||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db      1 GTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAG 60

Qy    240 aaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctgaagg 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAATAGATAAAATAAATGGAAAATTAGAAGGGTCTCCTGTTAAAGATGGTCTTCTGAAGG 120

Qy    300 ctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggacatgcaaa 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAA 180

Qy    360 ctttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaatgcaaa 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAA 240

Qy    420 agtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcagatcaga 479
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    241 AGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACATTGAGAGCAGATGAGA 300

Qy    480 tgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctgagagtc 539
          |  |||| |||||||| |||||||| |   |||||||| ||||||||||||||||||||
Db    301 TACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTC 360

Qy    540 tccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaaaagaaa 599
          || ||||||||||||||||||||||||||| |||||||||| | ||||||||||||||||
Db    361 TCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGGCTGCCCATCAAAAAGAAA 420

Qy    600 tggataaaataagtggaaaattagaagattcaact 634
          | ||| ||||| ||||||||||||||  ||  ||
Db    421 TAGATNAAATANATGGAAAATTAGAAGGGTCTCCT 455


RESULT 14
AX053808
LOCUS       AX053808      776 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 574 from Patent WO0073801.
ACCESSION   AX053808
VERSION     AX053808.1  GI:12228127
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 574 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      292 a    120 c    162 g    189 t     13 others
ORIGIN


  Query Match             48.8%;  Score 380;  DB 9;  Length 776;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-69;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    399 agcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatgaac 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 AGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAAC 64

Qy    459 aaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaa 518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 AAACATTGAGAGCAGATCAGATGTTCCCTTCAGAATCAAAACAAAAGAAGGTTGAAGAAA 124

Qy    519 attcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtaccca 578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 ATTCTTGGGATTCTGAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTGTACCCA 184

Qy    579 aggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaactagcc 638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 AGGCTACACATCAAAAAGAAATGGATAAAATAAGTGGAAAATTAGAAGATTCAACTAGCC 244

Qy    639 tatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatc 698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 TATCAAAAATCTTGGATACAGTTCATTCTTGTGAAAGAGCAAGGGAACTTCAAAAAGATC 304

Qy    699 actgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaa 758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    305 ACTGTGAACAACGTACAGGAAAAATGGAACAAATGAAAAAGAAGTTTTGTGTACTGAAAA 364

Qy    759 agaaactgtcagaagcaaaa 778
          ||||||||||||||||||||
Db    365 AGAAACTGTCAGAAGCAAAA 384


RESULT 15
AX053802
LOCUS       AX053802      771 bp    DNA             PAT       13-JAN-2001
DEFINITION  Sequence 568 from Patent WO0073801.
ACCESSION   AX053802
VERSION     AX053802.1  GI:12228121
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 568 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      251 a    145 c    155 g    216 t      4 others
ORIGIN
```

Query Match 43.0%; Score 334.8; DB 9; Length 771;
Best Local Similarity 94.1%; Pred. No. 3.4e-60;
Matches 348; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy    259 aaaattagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagt 318
          ||  ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 AATTTTCTTAGGGTCTCCTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGT 215

Qy    319 ttctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcc 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 TTCTATTCCAACTAAAGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCC 275

Qy    379 cgagaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagc 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 CGAGAAGCCATCTGCCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGC 335

Qy    439 cttggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaa 498
          |||||||||||||||||||||||||||||||||| ||||  |||| ||||||||  ||
Db    336 CTTGGAATTGAAGAATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAA 395

Qy    499 acaaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcaca 558
          |||||||  |   |||||||| |||||||||||||||||||||| ||||||||||||||||
Db    396 ACAAAAGGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACA 455

Qy    559 gaaggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaa 618
          |||||||||||||  ||||||||||  | ||||||||||||||  ||||||||||  ||||||
Db    456 GAAGGATGTGTGTTTACCCAAGGCTGCGCATCAAAAAGAAATAGATAAAATAAATGGAAA 515

Qy    619 attagaagat 628
          ||||||||  |
Db    516 ATTAGAAGGT 525
```

Search completed: April 30, 2001, 19:15:00
Job time: 37374 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 19:07:21 ; Search time 933.78 Seconds
(without alignments)
487.016 Million cell updates/sec

Title: US-09-721-183-5
Perfect score: 779
Sequence: 1 gtatacattctttattaatc.....gaaactgtcagaagcaaaaa 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	725.8	93.2	3681	22	F17975	Human breast cance
	2	723.8	92.9	1337	22	F17979	Human breast cance
	3	723.8	92.9	2307	22	F17980	Human breast cance
	4	722.2	92.7	2683	21	A59014	Nucleotide sequenc
	5	722.2	92.7	2683	21	A59015	Nucleotide sequenc
	6	622	79.8	1030	22	F22993	Human prostate can
	7	583.2	74.9	992	22	F23011	Human prostate can
	8	557.4	71.6	904	22	F22965	Human prostate can
	9	547.2	70.2	1729	22	F17978	Human breast cance
	10	530.2	68.1	664	22	F22980	Human prostate can
	11	512.2	65.8	814	22	F22967	Human prostate can
	12	507.8	65.2	820	22	F22970	Human prostate can
	13	503.8	64.7	1158	22	F22985	Human prostate can
	14	487.8	62.6	713	22	F22974	Human prostate can
	15	479.8	61.6	776	22	F22969	Human prostate can
	16	464.6	59.6	741	22	F22991	Human prostate can
	17	456.8	58.6	708	22	F22972	Human prostate can
	18	425.2	54.6	833	22	F22987	Human prostate can
	19	417.8	53.6	636	22	F22981	Human prostate can
	20	414.6	53.2	1665	22	F17692	Human breast cance
	21	380	48.8	776	22	F22995	Human prostate can
	22	334.8	43.0	771	22	F22989	Human prostate can
	23	318.4	40.9	766	22	F22986	Human prostate can
	24	274.8	35.3	278	21	A59008	Nucleotide sequenc
c	25	271	34.8	294	21	A59009	Nucleotide sequenc
	26	249.8	32.1	254	21	A59007	Nucleotide sequenc
	27	247	31.7	248	21	A59010	Nucleotide sequenc
	28	236.6	30.4	1424	22	F17976	Human breast cance
	29	236	30.3	609	22	F22997	Human prostate can
c	30	196.6	25.2	756	22	F22990	Human prostate can
	31	118	15.1	1681	22	F17693	Human breast cance
	32	81.8	10.5	1146	18	T66409	DNA encoding haema
	33	78	10.0	774	20	Z16196	Human gene express
	34	76.4	9.8	673	20	X40102	Gastric cancer ass
c	35	71.8	9.2	688	20	X40104	Gastric cancer ass
	36	66.8	8.6	469	20	V88140	EST clone DQ276.
	37	62.8	8.1	300	20	Z14520	Human gene express
c	38	60.4	7.8	714	20	X40103	Gastric cancer ass
c	39	58.6	7.5	180	21	C05602	Human secreted pro
	40	45.2	5.8	1047	20	X61805	B. burgdorferi ant
c	41	45.2	5.8	2892	20	X20292	Borrelia burgdorfe
	42	44.4	5.7	979	20	X61806	B. burgdorferi ant
	43	43.2	5.5	677	7	N60393	Sequence encoding
	44	43	5.5	1452	21	A70123	Plasmodium falcipa
	45	42.2	5.4	57728	22	C87588	Human 9p11 chromos

ALIGNMENTS

RESULT 1
F17975
ID F17975 standard; DNA; 3681 BP.
XX
AC F17975;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human breast cancer associated B726P consensus sequence #1.
XX
KW Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200060076-A2.
XX
PD 12-OCT-2000.
XX
PF 15-FEB-2000; 2000WO-US05308.
XX
PR 02-APR-1999; 99US-0285480.
PR 23-JUN-1999; 99US-0339338.
PR 02-SEP-1999; 99US-0389681.
PR 03-NOV-1999; 99US-0433826.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;
XX
DR WPI; 2001-122627/13.
DR P-PSDB; B50245.
XX
PT An isolated polypeptide useful for the treatment and diagnosis of
PT tumors e.g. breast cancer comprises at least an immunogenic portion of

PT a breast tumor protein -
XX
PS Claim 6; Page 225-227; 238pp; English.
XX
CC The present invention provides the coding sequences and some protein
CC sequences of proteins associated with breast cancer in humans. These
CC sequences can be used in the diagnosis and treatment of cancers,
CC particularly breast tumours.
XX
SQ Sequence 3681 BP; 1411 A; 680 C; 758 G; 826 T; 6 other;

Query Match 93.2%; Score 725.8; DB 22; Length 3681;
Best Local Similarity 98.3%; Pred. No. 2.1e-169;
Matches 730; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

```
Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1645 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 1704

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1705 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 1764

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1765 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 1824

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           ||||||||||:|:||||||||||||||||||||||||||||||||||||||| |||||
Db    1825 tacccaaggctrcrcatcaaaaagaaatagataaaataaatggaaaattagaagagtctc 1884

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||| ||| ||||||  ||||||||||  |||| ||||||||||||||||||||||||||
Db    1885 ctgataatgatggttttctgaaggctccctgcagaatgaaagtttctattccaactaaag 1944

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1945 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 2004

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2005 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 2064

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db    2065 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaasgttgaag 2124

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2125 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 2184

Qy     576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2185 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 2244

Qy     636 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2245 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 2304

Qy     696 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2305 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 2364

Qy     756 aaaagaaactgtcagaagcaaaa 778
           |||||||||||||||||||||||
Db    2365 aaaagaaactgtcagaagcaaaa 2387
```

RESULT 2
F17979
ID F17979 standard; cDNA; 1337 BP.
XX
AC F17979;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human breast cancer associated B726P splice sequence #4.
XX
KW Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200060076-A2.
XX
PD 12-OCT-2000.
XX
PF 15-FEB-2000; 2000WO-US05308.
XX
PR 02-APR-1999; 99US-0285480.
PR 23-JUN-1999; 99US-0339338.
PR 02-SEP-1999; 99US-0389681.
PR 03-NOV-1999; 99US-0433826.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;
XX
DR WPI; 2001-122627/13.
DR P-PSDB; B50259.
XX
PT An isolated polypeptide useful for the treatment and diagnosis of
PT tumors e.g. breast cancer comprises at least an immunogenic portion of
PT a breast tumor protein -
XX
PS Claim 6; Page 228-229; 238pp; English.
XX
CC The present invention provides the coding sequences and some protein
CC sequences of proteins associated with breast cancer in humans. These
CC sequences can be used in the diagnosis and treatment of cancers,
CC particularly breast tumours.
XX
SQ Sequence 1337 BP; 541 A; 213 C; 279 G; 304 T; 0 other;

Query Match 92.9%; Score 723.8; DB 22; Length 1337;
Best Local Similarity 98.4%; Pred. No. 4.9e-169;
Matches 731; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 219

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     220 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 279

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 339

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           |||||||||||  | ||||||||||||||||||||||||||||||||||||||| |||||
Db     340 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagagtctc 399

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||| ||| ||||||  ||||||||||  |||| ||||||||||||||||||||||||||
Db     400 ctgataatgatggttttctgaaggctccctgcagaatgaaagtttctattccaactaaag 459

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 519

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 579
```

```
Qy    456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 639

Qy    516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 699

Qy    576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 759

Qy    636 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 819

Qy    696 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    820 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 879

Qy    756 aaaagaaactgtcagaagcaaaa 778
          |||||||||||||||||||||||
Db    880 aaaagaaactgtcagaagcaaaa 902
```

RESULT 3
F17980
ID F17980 standard; cDNA; 2307 BP.
XX
AC F17980;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human breast cancer associated B726P splice sequence #5.
XX
KW Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200060076-A2.
XX
PD 12-OCT-2000.
XX
PF 15-FEB-2000; 2000WO-US05308.
XX
PR 02-APR-1999; 99US-0285480.
PR 23-JUN-1999; 99US-0339338.
PR 02-SEP-1999; 99US-0389681.
PR 03-NOV-1999; 99US-0433826.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;
XX
DR WPI; 2001-122627/13.
XX
PT An isolated polypeptide useful for the treatment and diagnosis of
PT tumors e.g. breast cancer comprises at least an immunogenic portion of
PT a breast tumor protein -
XX
PS Claim 6; Page 229-230; 238pp; English.
XX
CC The present invention provides the coding sequences and some protein
CC sequences of proteins associated with breast cancer in humans. These
CC sequences can be used in the diagnosis and treatment of cancers,
CC particularly breast tumours.
XX
SQ Sequence 2307 BP; 919 A; 409 C; 439 G; 540 T; 0 other;

Query Match 92.9%; Score 723.8; DB 22; Length 2307;
Best Local Similarity 98.4%; Pred. No. 5.8e-169;

Matches 731; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    288 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 347

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    348 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 407

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    408 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 467

Qy    216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    468 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagagtctc 527

Qy    276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
          ||| ||| |||||| |||||||| ||  |||| ||||||||||||||||||||||||||||
Db    528 ctgataatgatggttttctgaagtctccctgcagaatgaaagtttctattccaactaaag 587

Qy    336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    588 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 647

Qy    396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    648 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 707

Qy    456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    708 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaacgttgaag 767

Qy    516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    768 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 827

Qy    576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    828 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 887

Qy    636 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    888 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 947

Qy    696 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    948 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 1007

Qy    756 aaaagaaactgtcagaagcaaaa 778
          |||||||||||||||||||||||
Db   1008 aaaagaaactgtcagaagcaaaa 1030
```

RESULT 4
A59014
ID A59014 standard; cDNA; 2683 BP.
XX
AC A59014;
XX
DT 07-NOV-2000 (first entry)
XX
DE Nucleotide sequence of BS322 full length cDNA sequence.
XX
KW BS322; breast tissue marker; breast disease; breast cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200043420-A1.
XX
PD 27-JUL-2000.
XX

```
PF   21-JAN-2000; 2000WO-US01452.
XX
PR   21-JAN-1999;   99US-0234716.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI   Granados EN, Hodges SC, Klass MR, Kratochvil JD, Russell JC;
PI   Stroupe SD;
XX
DR   WPI; 2000-499217/44.
XX
PT   BS322-specific polynucleotides, polypeptides and antibodies, used in
PT   the detection and diagnosis of breast disease especially breast cancer
PT   -
XX
PS   Claim 12; Page 118-119; 126pp; English.
XX
CC   The present sequence represents the full length cDNA sequence
CC   encoding BS322 polypeptide. BS322 is a breast tissue marker. The
CC   BS322 polynucleotides and polypeptides are used to detect and
CC   diagnose breast disease, e.g. breast cancer. The BS322 polynucleotides
CC   are useful as a source of probes and primers, and the BS322 polypeptides
CC   are useful as antigens.
XX
SQ   Sequence 2683 BP; 1043 A; 462 C; 502 G; 676 T; 0 other;

  Query Match             92.7%;  Score 722.2;  DB 21;  Length 2683;
  Best Local Similarity   98.3%;  Pred. No. 1.5e-168;
  Matches  730;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     394 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 453

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     454 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 513

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     514 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 573

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     574 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagagtctc 633

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||| ||| |||||| ||||||||||  |||| ||||||||||||||||||||||||||||
Db     634 ctgataatgatggttttctgaaggctccctgcagaatgaaagtttctattccaactaaag 693

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     694 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 753

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     754 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 813

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     814 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaacgttgaag 873

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     874 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 933

Qy     576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     934 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 993

Qy     636 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 695
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     994 gcctatcaaaaatcttggatacaattcattcttgtgaaagagcaagggaacttcaaaaag 1053

Qy     696 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 755
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    1054 atcactgtgaacaatgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 1113

Qy     756 aaaagaaactgtcagaagcaaaa 778
           |||||||||||||||||||||||
Db    1114 aaaagaaactgtcagaagcaaaa 1136


RESULT   5
A59015
ID   A59015 standard; cDNA; 2683 BP.
XX
AC   A59015;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Nucleotide sequence of BS322 consensus cDNA sequence.
XX
KW   BS322; breast tissue marker; breast disease; breast cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             57..1253
FT                   /*tag=  a
FT                   /note= "encodes B07638"
FT   CDS             1172..2125
FT                   /*tag=  b
FT                   /note= "encodes B07639"
XX
PN   WO200043420-A1.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-US01452.
XX
PR   21-JAN-1999;   99US-0234716.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI   Granados EN, Hodges SC, Klass MR, Kratochvil JD, Russell JC;
PI   Stroupe SD;
XX
DR   WPI; 2000-499217/44.
DR   P-PSDB; B07638, B07639.
XX
PT   BS322-specific polynucleotides, polypeptides and antibodies, used in
PT   the detection and diagnosis of breast disease especially breast cancer
PT   -
XX
PS   Claim 12; Page 119-120; 126pp; English.
XX
CC   The present sequence encodes a human BS322 polypeptide. BS322 is
CC   a breast tissue marker. The BS322 polynucleotides and polypeptides
CC   are used to detect and diagnose breast disease, e.g. breast cancer.
CC   The BS322 polynucleotides are useful as a source of probes and
CC   primers, and the BS322 polypeptides are useful as antigens.
XX
SQ   Sequence 2683 BP; 1043 A; 462 C; 502 G; 676 T; 0 other;


  Query Match             92.7%;  Score 722.2;  DB 21;  Length 2683;
  Best Local Similarity   98.3%;  Pred. No. 1.5e-168;
  Matches  730;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     394 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 453
```

```
Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    454 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 513

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    514 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 573

Qy    216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    574 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagagtctc 633

Qy    276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
          ||| ||| |||||| |||||||||||  |||| ||||||||||||||||||||||||||
Db    634 ctgataatgatggttttctgaaggctccctgcagaatgaaagtttctattccaactaaag 693

Qy    336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    694 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 753

Qy    396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    754 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 813

Qy    456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    814 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaacgttgaag 873

Qy    516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    874 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 933

Qy    576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    934 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaacta 993

Qy    636 gcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaaag 695
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    994 gcctatcaaaaatcttggatacaattcattcttgtgaaagagcaagggaacttcaaaaag 1053

Qy    696 atcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 755
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1054 atcactgtgaacaatgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactga 1113

Qy    756 aaaagaaactgtcagaagcaaaa 778
          |||||||||||||||||||||||
Db   1114 aaaagaaactgtcagaagcaaaa 1136
```

RESULT 6
F22993
ID F22993 standard; cDNA; 1030 BP.
XX
AC F22993;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:572.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Claim 50; Page 459-460; 799pp; English.
XX
CC F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC nucleotide sequences encoding human breast, gastric and prostate cancer
CC associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC activity and can be used in the production of cancer vaccines. The human
CC CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC useful for diagnosing and treating a condition characterised by
CC expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ Sequence 1030 BP; 383 A; 170 C; 219 G; 258 T; 0 other;

Query Match 79.8%; Score 622; DB 22; Length 1030;
Best Local Similarity 98.4%; Pred. No. 5.1e-144;
Matches 628; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy    141 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 60

Qy    201 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 120

Qy    261 aattagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagttt 320
          ||||||||| |||||||| ||| |||||| |||||||||||  |||| ||||||||||||
Db    121 aattagaagagtctcctgataatgatggttttctgaaggctccctgcagaatgaaagttt 180

Qy    321 ctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccg 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccg 240

Qy    381 agaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagcct 440
          |||||||||||||| |  ||||||||||||||||||||||||||||||||||||||||||
Db    241 agaagccatctgccatttagcctgccattgaaatgcaaaagtctgttccaaataaagcct 300

Qy    441 tggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaac 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 tggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaac 360

Qy    501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 420

Qy    561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 480

Qy    621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 540

Qy    681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 600

Qy    741 agttttgtgtactgaaaaagaaactgtcagaagcaaaa 778
          |||||||||||||||||||||||||||||||||||||
```

```
Db    601 agttttgtgtactgaaaaagaaactgtcagaagcaaaa 638

RESULT   7
F23011
ID   F23011 standard; cDNA; 992 BP.
XX
AC   F23011;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human prostate cancer associated antigen nucleotide sequence SEQ ID:590.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 465; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 992 BP; 348 A; 163 C; 200 G; 243 T; 38 other;

  Query Match           74.9%;  Score 583.2;  DB 22;  Length 992;
  Best Local Similarity  94.7%;  Pred. No. 1.8e-134;
  Matches  605;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy    141 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aacaaaaggactatgaagaaagttcttgggattctgagagtctctgtgagactgtttcac 60

Qy    201 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agaaggatgtgtgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaa 120

Qy    261 aattagaagggtctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagttt 320
          ||||| ||| |||||||| ||| |||||| |||||||||||  ||||  ||||||||||||
Db    121 aattanaagagtctcctgataatgatggttttctgaaggctccctgcanaatgaaagttt 180

Qy    321 ctattccaactaaagccttagaattgatggacatgcaaactttcaaagcagagcctcccg 380
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    181 ctattccaactaaagccttanaattgatggacatgcaaactttcaaagcagagcctcccg 240

Qy    381 agaagccatctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagcct 440
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 agaagccntctgccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagcct 300

Qy    441 tggaattgaagaatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaac 500
          |||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db    301 tggaattgaagaatgaacnaacattgagagcagatcagatgttcccttcanaatcaaaac 360

Qy    501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
          |||||||| |||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db    361 aaaagaaggntgaagaaaattcttgggattctganagtctccntgagactgtttcacaga 420

Qy    561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 480

Qy    621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    481 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagancaa 540

Qy    681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
          |||||||| ||||||||      |  | |||||  ||||||||||||||||||||||||
Db    541 nggaacttcnaaaagatcctgtgaacaaccgtacnngaaaaatggaacaaatgaaaaaga 600

Qy    741 agtttt-gtgtactgaaaaagaaactgtcagaagcaaaa 778
          |||||| |||| ||||||||||||||||| ||||| |||
Db    601 agttttggtgtnctgaaaaagaaactgtccgaagccaaa 639

RESULT   8
F22965
ID   F22965 standard; cDNA; 904 BP.
XX
AC   F22965;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human prostate cancer associated antigen nucleotide sequence SEQ ID:544.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 450-451; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
```

```
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 904 BP; 315 A; 173 C; 200 G; 216 T; 0 other;


  Query Match             71.6%;  Score 557.4;  DB 22;  Length 904;
  Best Local Similarity   95.7%;  Pred. No. 3.8e-128;
  Matches  573;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      46 ccttcgagcctgccattgatatgcaaaagtctgttccaaataaagccttggaattgaaga 105

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     106 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 165

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     166 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 225

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaaatggaaaattagaagggtctc 275
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db     226 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 285

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     286 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 345

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 405

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     406 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 465

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           |||||||||||||||||| |||| |||| |||||||| ||||||||  |   |||||
Db     466 aacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaag 525

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||| |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db     526 aaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtttac 585

Qy     576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattcaact 634
           |||||||| | |||||||||||||| |||||||||| ||||||||||||||  ||  ||
Db     586 ccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtctcct 644


RESULT   9
F17978
ID   F17978 standard; cDNA; 1729 BP.
XX
AC   F17978;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human breast cancer associated B726P splice sequence #3.
XX
KW   Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200060076-A2.
XX
PD   12-OCT-2000.
XX
PF   15-FEB-2000; 2000WO-US05308.
XX
PR   02-APR-1999;   99US-0285480.
PR   23-JUN-1999;   99US-0339338.
PR   02-SEP-1999;   99US-0389681.
PR   03-NOV-1999;   99US-0433826.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Dillon DC,  Mitcham JL,  Xu J,  Harlocker SL;
XX
DR   WPI; 2001-122627/13.
DR   P-PSDB; B50258.
XX
PT   An isolated polypeptide useful for the treatment and diagnosis of
PT   tumors e.g. breast cancer comprises at least an immunogenic portion of
PT   a breast tumor protein -
XX
PS   Claim 6; Page 228; 238pp; English.
XX
CC   The present invention provides the coding sequences and some protein
CC   sequences of proteins associated with breast cancer in humans. These
CC   sequences can be used in the diagnosis and treatment of cancers,
CC   particularly breast tumours.
XX
SQ   Sequence 1729 BP; 669 A; 296 C; 371 G; 391 T; 2 other;


  Query Match             70.2%;  Score 547.2;  DB 22;  Length 1729;
  Best Local Similarity   95.1%;  Pred. No. 1.5e-125;
  Matches  564;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||| ||||||||||||||||||| |||||||||||||||||||||| |
Db     880 ccttcgagcctgccactgaaatgcaaaagtctgtcccaaataaagccttggaattgaaaa 939

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 999

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           |||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1000 aagaaaattcttgggatactgagagtctctgtgagactgtttcacagaaggatgtgtgtt 1059

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db    1060 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 1119

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 ctggtaaanatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 1179

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 1239

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 1299

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           |||||||||||||||||| ||||  |||| |||||||| |||||||| |   |||||
Db    1300 aacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaag 1359

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||| |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db    1360 aaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtttac 1419

Qy     576 ccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagat 628
           |||||||| | |||||||||||||| |||||||||| |||||||||||||| |
Db    1420 ccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaaggt 1472


RESULT  10
F22980
ID   F22980 standard; cDNA; 664 BP.
XX
```

```
AC   F22980;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human prostate cancer associated antigen nucleotide sequence SEQ ID:559.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 455; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 664 BP; 236 A; 115 C; 144 G; 166 T; 3 other;

  Query Match             68.1%;  Score 530.2;  DB 22;  Length 664;
  Best Local Similarity   95.3%;  Pred. No. 1.7e-121;
  Matches  566;  Conservative    0;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           ||  |  |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       4 ccttcgagcctgccattgatatgcaaaagtctgttccaaataaagccttggaattgaaga 63

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 123

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 183

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           |||||||||||  | ||||||||||||||||||||||||||||||||||||||||||||
Db     184 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 243

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 303

Qy     336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 363

Qy     396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 423

Qy     456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
           ||||||||||||||||||| ||||  |||| ||||||||| |||||||||  |   |||||
Db     424 aacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaag 483

Qy     516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
           ||| |||||||||||||||||||||| |||||||||||| |||||||||||||| |||
Db     484 aaagttcttgggattctgagagtctctgtgagactgtttc-cagaaggatgtgtgtttac 542

Qy     576 ccaaggc-tacacatcaaaaagaaatggataaaataagtggaaaattagaagat 628
           |||| || | | | ||||||||||||| |||||||||| ||||||||||||  |
Db     543 ccaangcttgcgcntcaaaaagaaatagataaaataaatggaaaattagaangt 596


RESULT  11
F22967
ID   F22967 standard; cDNA; 814 BP.
XX
AC   F22967;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human prostate cancer associated antigen nucleotide sequence SEQ ID:546.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 451; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 814 BP; 278 A; 159 C; 173 G; 195 T; 9 other;

  Query Match             65.8%;  Score 512.2;  DB 22;  Length 814;
  Best Local Similarity   93.7%;  Pred. No. 4.9e-117;
  Matches  554;  Conservative    0;  Mismatches   35;  Indels    2;  Gaps    2;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
```

```
Db     31 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 90

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 150

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 210

Qy    216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
          |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db    211 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 270

Qy    276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaag 330

Qy    336 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcct 395
          ||||||||||||||||||||||||||||||||||||||||||||||||||      ||
Db    331 ccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagcccatctgct 390

Qy    396 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391 tcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatg 450

Qy    456 aacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaag 515
          |||||||||||||||||| |||| |||| |||||||| ||||||||| |   |||||
Db    451 aacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaag 510

Qy    516 aaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtac 575
          ||| |||||||||||||||||||| |||||||||||| || |||||||||||| | |
Db    511 aaagttcttgggattctgagagtctctgtgagactgtttc-canaaggatgtgtgt-ttc 568

Qy    576 ccaaggctacacatcaaaaagaaatggatcaaaataagtggaaaattagaag 626
          |||||||| | |||||||||||||| ||| ||||||  ||||| |   | |
Db    569 ccaaggctgcgcatcaaaaagaaatagattaaataatgggaaattagaang 619
```

RESULT 12
F22970
ID F22970 standard; cDNA; 820 BP.
XX
AC F22970;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:549.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Claim 50; Page 452; 799pp; English.
XX
CC F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC nucleotide sequences encoding human breast, gastric and prostate cancer
CC associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC activity and can be used in the production of cancer vaccines. The human
CC CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC useful for diagnosing and treating a condition characterised by
CC expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ Sequence 820 BP; 282 A; 155 C; 168 G; 203 T; 12 other;

Query Match 65.2%; Score 507.8; DB 22; Length 820;
Best Local Similarity 98.5%; Pred. No. 5.9e-116;
Matches 523; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

```
Qy     94 gaatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggacta 153
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gaatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggacta 60

Qy    154 tgaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtg 213
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tgaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtg 120

Qy    214 tttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtc 273
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    121 tttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagagtc 180

Qy    274 tcctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaa 333
          ||||| ||| |||||| |||||||||||  |||| ||||||||||||||||||||||||
Db    181 tcctgataatgatggttttctgaaggctccctgcagaatgaaagtttctattccaactaa 240

Qy    334 agccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgc 393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 agccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgc 300

Qy    394 cttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaa 453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 cttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaa 360

Qy    454 tgaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttga 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tgaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttga 420

Qy    514 agaaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgt 573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 agaaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgt 480

Qy    574 acccaaggctacacatcaaaaagaaatggataaaataagtggaaaattaga 624
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 -cccaaggctacacatcaaaaagaaatggataaaataagtggaaaattaga 530
```

RESULT 13
F22985
ID F22985 standard; cDNA; 1158 BP.
XX
AC F22985;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:564.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX

```
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 457; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 1158 BP; 399 A; 216 C; 241 G; 300 T; 2 other;
```

```
  Query Match             64.7%;  Score 503.8;  DB 22;  Length 1158;
  Best Local Similarity   89.9%;  Pred. No. 6.3e-115;
  Matches  569;  Conservative    0;  Mismatches   22;  Indels   42;  Gaps    1;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 ccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 301

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 361

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 421

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaag------ 269
           |||||||||||| | ||||||||||||||||||||||||||||||||||||||      
Db     422 tacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaaggtagat 481

Qy     270 ------------------------------------ggtctcctgttaaagatggtcttc 293
                                                ||||||||||||||||||||||||
Db     482 atgctgctgaatttagaacattctctgcaatgataaggtctcctgttaaagatggtcttc 541

Qy     294 tgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggaca 353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 tgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggaca 601

Qy     354 tgcaaactttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaa 413
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     602 tgcaaactttcaaagcagaacctcccgagaagccatctgccttcgagcctgccattgaaa 661

Qy     414 tgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcag 473
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 tgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcag 721

Qy     474 atcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctg 533
           || ||||  |||| |||||||| ||||||||| |   |||||||| ||||||||||||||
Db     722 atgagatactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctg 781

Qy     534 agagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaa 593
           ||||||||| ||||||||||||||||||||||||||||| |||||||||| | |||||||
Db     782 agagtctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctgcgcatcaaa 841

Qy     594 aagaaatggataaaataagtggaaaattagaag 626
           ||||||| |||||||||| ||||||||||||||
Db     842 aagaaatagataaaataaatggaaaattagaag 874
```

```
RESULT  14
F22974
ID   F22974 standard; cDNA; 713 BP.
XX
AC   F22974;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human prostate cancer associated antigen nucleotide sequence SEQ ID:553.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids  encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 453-454; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 713 BP; 250 A; 133 C; 153 G; 172 T; 5 other;
```

```
  Query Match             62.6%;  Score 487.8;  DB 22;  Length 713;
  Best Local Similarity   95.5%;  Pred. No. 4.8e-111;
  Matches  512;  Conservative    0;  Mismatches   23;  Indels    1;  Gaps    1;

Qy      95 aatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactat 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 aatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactat 60

Qy     155 gaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgt 214
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgt 120
```

```
Qy    215 ttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtct 274
          |||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db    121 ttacccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattagaagggtct 180

Qy    275 cctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaa 334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cctgttaaagatggtctttctgaaggctaactgcggaatgaaagtttctattccaactaaa 240

Qy    335 gccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcc 394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgcc 300

Qy    395 ttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaat 454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaat 360

Qy    455 gaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaa 514
          |||||||||||||||||||| |||| |||| |||||||| |||||||| | . ||||
Db    361 gaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaa 420

Qy    515 gaaaattcttgggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgta 574
          |||| |||||||||||||||||||| |||||||||||||| |||||||||||| | |
Db    421 gaaagttcttgggattctgagagtctctgtgagactgtttcacanaaggatgtgtgt-tt 479

Qy    575 cccaaggctacacatcaaaaagaaatggataaaataagtggaaaattagaagattc 630
          ||||||||| | |||||||||||||| |||||||||| | |  ||
Db    480 cccaaggctgcgcatcaaaaagaaatagataaaataaatggaaaattaaagggctc 535
```

RESULT 15
F22969
ID F22969 standard; cDNA; 776 BP.
XX
AC F22969;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human prostate cancer associated antigen nucleotide sequence SEQ ID:548.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Claim 50; Page 452; 799pp; English.
XX
CC F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC nucleotide sequences encoding human breast, gastric and prostate cancer
CC associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC activity and can be used in the production of cancer vaccines. The human
CC CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC useful for diagnosing and treating a condition characterised by
CC expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ Sequence 776 BP; 273 A; 142 C; 167 G; 188 T; 6 other;

Query Match 61.6%; Score 479.8; DB 22; Length 776;
Best Local Similarity 95.7%; Pred. No. 4.5e-109;
Matches 493; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy    120 agatactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgaga 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 agatactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgaga 60

Qy    180 gtctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctacacatcaaaaag 239
          |||||||||||||||||||||||||||||||||||||||||||| | ||||||||||
Db     61 gtctctgtgagactgtttcacagaaggatgtgtgtttacccaaggctgcgcatcaaaaag 120

Qy    240 aaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctgaagg 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctgaagg 180

Qy    300 ctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggacatgcaaa 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggacatgcaaa 240

Qy    360 ctttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaatgcaaa 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ctttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaatgcaaa 300

Qy    420 agtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcagatcaga 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    301 agtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcagatgaga 360

Qy    480 tgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctgagagtc 539
          |  |||| |||||||| |||||||| |   |||||||| ||||||||||||||||||||
Db    361 tactcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgagagtc 420

Qy    540 tccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaaaagaaa 599
          || ||||||||||||||||||||||||||||| |||||||||||| | ||||||||||||
Db    421 tctgtgagactgtttcacagaaggatgtgtgtttacccaaggctgcgcatcaaaaagaaa 480

Qy    600 tggataaaataagtggaaaattagaagattcaact 634
          | |||||||||| |||||||||||||  ||  ||
Db    481 tagataaaataaatggaaaattagaagggtctcct 515
```

Search completed: April 30, 2001, 19:07:29
Job time: 36863 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 00:33:32 ; Search time 330.07 Seconds
(without alignments)
412.081 Million cell updates/sec

Title: US-09-721-183-5
Perfect score: 779
Sequence: 1 gtatacattctttattaatc.....gaaactgtcagaagcaaaaa 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	81.8	10.5	1146	1	US-08-487-810-1	Sequence 1, Appli
c 2	66	8.5	7218	1	US-08-232-463-14	Sequence 14, Appl
3	39.2	5.0	2319	2	US-08-966-389-5	Sequence 5, Appli
4	39.2	5.0	2319	2	US-09-103-509-5	Sequence 5, Appli
5	39.2	5.0	2319	2	US-09-102-644-5	Sequence 5, Appli
6	39.2	5.0	2319	2	US-09-218-032-5	Sequence 5, Appli
7	39.2	5.0	3345	2	US-08-966-389-8	Sequence 8, Appli
8	39.2	5.0	3345	2	US-09-103-509-8	Sequence 8, Appli
9	39.2	5.0	3345	2	US-09-102-644-8	Sequence 8, Appli
10	39.2	5.0	3345	2	US-09-218-032-8	Sequence 8, Appli
11	38.8	5.0	4849	2	US-08-540-804-13	Sequence 13, Appl
12	38.8	5.0	4849	2	US-08-218-265-13	Sequence 13, Appl
13	38.8	5.0	4849	3	US-08-521-872-13	Sequence 13, Appl
14	38.2	4.9	1736	4	US-09-182-816-22	Sequence 22, Appl
c 15	38.2	4.9	1736	4	US-09-182-816-24	Sequence 24, Appl
16	38.2	4.9	1736	4	US-09-471-528-22	Sequence 22, Appl
c 17	38.2	4.9	1736	4	US-09-471-528-24	Sequence 24, Appl
c 18	38.2	4.9	7587	4	US-08-378-313-22	Sequence 22, Appl
19	38	4.9	2223	1	US-08-257-073-4	Sequence 4, Appli
20	37.6	4.8	3600	3	US-08-894-731-1	Sequence 1, Appli
21	35.2	4.5	1498	1	US-08-118-469A-1	Sequence 1, Appli
22	35.2	4.5	1498	1	US-08-909-119-1	Sequence 1, Appli
23	34.8	4.5	240	1	US-08-628-417-6	Sequence 6, Appli
24	34.6	4.4	5181	1	US-08-257-073-10	Sequence 10, Appl
25	34.4	4.4	1886	6	5210183-1	Patent No. 5210183
26	34.4	4.4	6243	2	US-09-056-075-1	Sequence 1, Appli
c 27	34.2	4.4	1519	1	US-07-971-759-19	Sequence 19, Appl
28	34	4.4	2447	2	US-09-014-969-14	Sequence 14, Appl
29	34	4.4	5761	1	US-07-749-001-2	Sequence 2, Appli
30	34	4.4	5761	1	US-08-154-198-2	Sequence 2, Appli
31	34	4.4	5761	1	US-08-463-335-2	Sequence 2, Appli
32	34	4.4	5761	2	US-08-464-023A-2	Sequence 2, Appli
33	34	4.4	56516	2	US-08-996-306-1	Sequence 1, Appli
34	33.6	4.3	7101	1	US-08-480-604A-9	Sequence 9, Appli
35	33.6	4.3	7101	2	US-08-405-496A-9	Sequence 9, Appli
36	33.2	4.3	1415	1	US-08-413-118-126	Sequence 126, App
37	33.2	4.3	1415	3	US-08-473-446-126	Sequence 126, App
38	33.2	4.3	5361	4	US-08-973-462-2	Sequence 2, Appli
39	33.2	4.3	6152	4	US-08-973-462-1	Sequence 1, Appli
40	33	4.2	5177	6	5352450-1	Patent No. 5352450
41	32.8	4.2	8585	1	US-08-030-096-3	Sequence 3, Appli
42	32.6	4.2	482	2	US-08-544-822-12	Sequence 12, Appl
43	32.6	4.2	482	3	US-09-070-964-12	Sequence 12, Appl
44	32.6	4.2	535	2	US-07-635-196A-2	Sequence 2, Appli
45	32.6	4.2	763	2	US-08-544-822-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-487-810-1
; Sequence 1, Application US/08487810
; Patent No. 5618695
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA ENCODING HEM-1, A GENE EXPRESSED BY
; TITLE OF INVENTION: SCLEROSING HEMANGIOMA CELLS
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-487-810-1

Query Match 10.5%; Score 81.8; DB 1; Length 1146;
Best Local Similarity 51.6%; Pred. No. 7.2e-13;
Matches 311; Conservative 0; Mismatches 232; Indels 60; Gaps 3;

```
Qy    236 aaagaaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctg 295
          | ||||||| |  |   | |   |         || |||||||   ||| |       | |
Db    446 AGAGAAATAAAGGATGGAGAAAAATCTGGGACAGTGTCTCCTCAGAAACAATCGGCCCAG 505

Qy    296 aaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatggacatg 355
          |||| ||  |    || |||||||||| ||   | ||  |||  |  | | | || |
Db    506 AAGGTTATATTTAAAAAGAAAGTTTCTCTTTTGAATATTGCCACAAGAATAACGGGCGGT 565

Qy    356 caaactttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattgaaatg 415
            |  |    | |||||  ||| ||||| |   |  ||||  || || | |||||||
Db    566 TGGAAATCTGGAACAGAGTATCCTGAGAATCTGCCCACCTTGAAGGCTACAATTGAAAAT 625

Qy    416 caaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagcagat 475
           |||| |||||||  |||| ||||    || ||||  ||| ||||||||  | | ||||
Db    626 AAAAATTCTGTTCTGAATACAGCCACCAAAATGAAAGATGTACAAACATCCACACCAGAA 685

Qy    476 cagatgt------tcccttcagaatcaaaacaaaagaaggttgaagaaaat---tcttgg 526
          ||    |      |  | |||||   | | |||||||  | ||||||||  ||     |
Db    686 CAAGACTTAGAAATGGCATCAGAGGGAGAGCAAAAGAGGCTTGAAGAATATGAAAATAAC 745
```

```
Qy    527 gattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctaca 586
          |  |  ||          | |    |||   | | ||  |   ||     |      ||
Db    746 CAGCCACAGGTGAAAAACCAAATACATTCTAGGGATGACCTTGATGACATAATTCAGTCA 805

Qy    587 catcaaaaag---------------------------------------------------- 596
            |||||  ||
Db    806 TCTCAAACAGTCTCAGAGGACGGTGACTCGCTTTGCTGTAATTGTAAGAATGTCATATTA 865

Qy    597 -aaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatcttggat 655
           || ||| ||          || |   ||||||    |  |||||  |||||      |
Db    866 CTCATTGATCAACATGAAATGAAGTGTAAAGATTGTGTTCACCTATTGAAAATTAAAAAG 925

Qy    656 acagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaacgtaca 715
          ||| ||  ||  ||| |||||  ||  |||||| || |  |||||||||| |||| || |
Db    926 ACATTTTGTTTATGTAAAAGATTAACAGAACTTAAAGATAATCACTGTGAGCAACTTAGA 985

Qy    716 ggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgtcagaagca 775
          | ||||||   | || |||||| |||  | |||||||  |||||| ||| || |||  |
Db    986 GTAAAAATTCGAAAACTGAAAAATAAGGCTAGTGTACTACAAAAGAGACTATCTGAAAAA 1045

Qy    776 aaa 778
            ||
Db   1046 GAA 1048
```

```
RESULT   2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             8.5%;  Score 66;  DB 1;  Length 7218;
  Best Local Similarity   3.2%;  Pred. No. 2.1e-08;
  Matches   12;  Conservative  224;  Mismatches  134;  Indels    0;  Gaps    0;

Qy    410 gaaatgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgaga 469
          || ||  || | | || | |::: :::    ::::  :::::: ::: ::: :  ::::
Db   1454 GAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

Qy    470 gcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggat 529
          : :::  ::: :         ::::  :::: :::::::::  ::::::::     ::::
Db   1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

Qy    530 tctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacat 589
            :::::     : ::::  :    : ::::::: : : : : :   ::::  : : :
Db   1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

Qy    590 caaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatc 649
            ::::::::: ::: :::: ::: :::::  ::::::   ::  ::   :  :::::
Db   1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

Qy    650 ttggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaa 709
            ::: : ::   :     : ::::::: :::::::    :::::::  :  : ::: ::
Db   1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

Qy    710 cgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgtca 769
          : : :::::::: :::: ::: ::::::::::    : : :  ::::::::::  :  : :
Db   1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

Qy    770 gaagcaaaaa 779
          :::: :::::
Db   1094 RRRRRRRRRR 1085
```

```
RESULT   3
US-08-966-389-5
; Sequence 5, Application US/08966389
; Patent No. 5843748
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,389
;      FILING DATE:  07-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-966-389-5


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 2319;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy    591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
          || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    743 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 802

Qy    651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
          || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |      |
Db    803 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 862

Qy    711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
            | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db    863 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 922

Qy    768 cagaagca 775
            |||| |
Db    923 AGGAAGAA 930


RESULT   4
US-09-103-509-5
; Sequence 5, Application US/09103509
; Patent No. 5876975
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,509
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:  07-NOV-1997
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-103-509-5


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 2319;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy    591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
          || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    743 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 802

Qy    651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
          || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |      |
Db    803 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 862

Qy    711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
            | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db    863 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 922

Qy    768 cagaagca 775
            |||| |
Db    923 AGGAAGAA 930


RESULT   5
US-09-102-644-5
; Sequence 5, Application US/09102644
; Patent No. 5910436
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,644
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:  07-NOV-1997
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-102-644-5


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 2319;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db     743 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 802

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db     803 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 862

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
             | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db     863 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 922

Qy     768 cagaagca 775
             |||| |
Db     923 AGGAAGAA 930


RESULT   6
US-09-218-032-5
; Sequence 5, Application US/09218032
; Patent No. 5993889
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/218,032
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-218-032-5


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 2319;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db     743 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 802

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db     803 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 862

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
             | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db     863 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 922

Qy     768 cagaagca 775
             |||| |
Db     923 AGGAAGAA 930


RESULT   7
US-08-966-389-8
; Sequence 8, Application US/08966389
; Patent No. 5843748
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,389
;      FILING DATE:  07-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  8:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermoanaerobium brockii
;      STRAIN:  ATCC 35047
;    FEATURE:
;    FEATURE:  1-595 5'-UTR
;    FEATURE:  E
;    FEATURE:  596-2917 mat peptide
;    FEATURE:  S
;    FEATURE:  2918-3345 3'-UTR
;    FEATURE:  E
US-08-966-389-8


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 3345;
  Best Local Similarity   54.3%;  Pred. No. 0.19;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    1341 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 1400

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db    1401 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 1460

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
           | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db    1461 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 1520

Qy     768 cagaagca 775
             |||| |
Db    1521 AGGAAGAA 1528


RESULT   8
US-09-103-509-8
; Sequence 8, Application US/09103509
; Patent No. 5876975
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,509
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:  07-NOV-1997
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA-6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermoanaerobium brockii
;      STRAIN:  ATCC 35047
;    FEATURE:
;    FEATURE:  1-595 5'-UTR
;    FEATURE:  E
;    FEATURE:  596-2917 mat peptide
;    FEATURE:  S
;    FEATURE:  2918-3345 3'-UTR
;    FEATURE:  E
US-09-103-509-8


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 3345;
  Best Local Similarity   54.3%;  Pred. No. 0.19;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    1341 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 1400

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db    1401 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 1460

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
           | |   || | |||| ||| ||   |||||  |    | || |  || | | |||  |
Db    1461 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 1520

Qy     768 cagaagca 775
             |||| |
Db    1521 AGGAAGAA 1528


RESULT   9
US-09-102-644-8
; Sequence 8, Application US/09102644
; Patent No. 5910436
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,644
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:  07-NOV-1997
;      APPLICATION NUMBER:  JP 311,232/1996
;      FILING DATE:  8-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermoanaerobium brockii
;      STRAIN:  ATCC 35047
;    FEATURE:
;    FEATURE:  1-595 5'-UTR
;    FEATURE:  E
;    FEATURE:  596-2917 mat peptide
;    FEATURE:  S
;    FEATURE:  2918-3345 3'-UTR
;    FEATURE:  E
US-09-102-644-8


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 3345;
  Best Local Similarity   54.3%;  Pred. No. 0.19;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    1341 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 1400

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db    1401 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 1460

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
           | |   || | |||| ||| ||   |||||  |   | || |  || | | |||  |
Db    1461 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 1520

Qy     768 cagaagca 775
             |||| |
Db    1521 AGGAAGAA 1528


RESULT  10
US-09-218-032-8
; Sequence 8, Application US/09218032
; Patent No. 5993889
;  GENERAL INFORMATION:
;    APPLICANT:  Tetsuya NAKADA
;    APPLICANT:  Michio KUBOTA
;    APPLICANT:  Hiroto CHAEN
;    APPLICANT:  Toshio MIYAKE
;    TITLE OF INVENTION:  TREHALOSE PHOSPHORYLASE, ITS PREPARATION
;    TITLE OF INVENTION:  AND USES
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/218,032
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,389
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 61,716/1997
;      FILING DATE:  3-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  NAKADA=6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermoanaerobium brockii
;      STRAIN:  ATCC 35047
;    FEATURE:
;    FEATURE:  1-595 5'-UTR
;    FEATURE:  E
;    FEATURE:  596-2917 mat peptide
;    FEATURE:  S
;    FEATURE:  2918-3345 3'-UTR
;    FEATURE:  E
US-09-218-032-8


  Query Match             5.0%;  Score 39.2;  DB 2;  Length 3345;
  Best Local Similarity   54.3%;  Pred. No. 0.19;
  Matches  102;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy     591 aaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatct 650
           || ||||    |  ||| ||| || | |  |||| | |  ||   ||   ||| | |  |
Db    1341 AAGAAGATGGAGTAAAAGTAATTGTAGATCTAGAGGCTGAAAAAGGCACCTCATACACTT 1400

Qy     651 tggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaac 710
           || ||| | ||  ||| | | | |   ||| |||  || |  || || | |  |     |
Db    1401 TGAATAAATTTATTTCCTATTACACTTCAAAGGATTTTGATGAAAATAAATTGGTTGCTC 1460

Qy     711 gtac---aggaaaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgt 767
           | |   || | |||| ||| ||   |||||  |   | || |  || | | |||  |
Db    1461 TTGCTTTAGAAGAAATAGAAAAAGCCAAAAATGACGGCTTTGAAACGATAGAAAAAGAGC 1520

Qy     768 cagaagca 775
             |||| |
Db    1521 AGGAAGAA 1528


RESULT  11
US-08-540-804-13
```

```
; Sequence 13, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
;   APPLICANT: Young, Richard A.
;   APPLICANT: Koleske, Anthony J.
;   APPLICANT: Thompson, Craig M.
;   APPLICANT: Chao, David M.
;   TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
;   TITLE OF INVENTION: Transcription and Methods of Use Therefor
;   NUMBER OF SEQUENCES: 39
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: Massachusetts
;     COUNTRY: USA
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/540,804
;     FILING DATE: 11-OCT-1995
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/521,872
;     FILING DATE: 21-AUG-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/218,265
;     FILING DATE: 25-MAR-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: WHI94-03A2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 4849 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 148..4407
US-08-540-804-13


  Query Match             5.0%;  Score 38.8;  DB 2;  Length 4849;
  Best Local Similarity   47.5%;  Pred. No. 0.27;
  Matches  115;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy     501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
           | || ||  |  | || |||   |   |  ||| ||     |       || || ||| |
Db    1349 ATAATAACATCAATGAGAATGAATTATTTAATGATAGGAAACAAACTACAGTATCAAATG 1408

Qy     561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
           |    | |       ||   |   ||  | |   | | ||  | || |  |   | |||
Db    1409 ACTTAGAGAACAGCCCACTGAAAACGGAACTGGAGGCAAATGGTAGATCACTCGAAAAGG 1468

Qy     621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
           || |  ||||    |||    ||  || | | || ||| |||||  |   ||  || ||
Db    1469 TAAATAATTCCGTGAGCAAGACAGGAAGCGTAGACACACTTCATAATAAAGAGGGAACAC 1528

Qy     681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
            |||||  | | || |  ||   | |  |  |  |  |||| ||    |||  | | |
Db    1529 TGGAACAACGAGAACAGAACGAAAATCTGCCAAGTGATAAAAGTGACTCTATGGTAGACA 1588

Qy     741 ag 742
           ||
Db    1589 AG 1590


RESULT 12
US-08-218-265-13
; Sequence 13, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
;   APPLICANT: Young, Richard A.
;   APPLICANT: Koleske, Anthony J.
;   APPLICANT: Thompson, Craig M.
;   TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
;   TITLE OF INVENTION: Transcription and Methods of Use Thereof
;   NUMBER OF SEQUENCES: 35
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: MA
;     COUNTRY: US
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/218,265
;     FILING DATE: 25-MAR-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: WHI94-03
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 4849 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 148..4407
US-08-218-265-13


  Query Match             5.0%;  Score 38.8;  DB 2;  Length 4849;
  Best Local Similarity   47.5%;  Pred. No. 0.27;
  Matches  115;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy     501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
           | || ||  |  | || |||   |   |  ||| ||     |       || || ||| |
Db    1349 ATAATAACATCAATGAGAATGAATTATTTAATGATAGGAAACAAACTACAGTATCAAATG 1408

Qy     561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
           |    | |       ||   |   ||  | |   | | ||  | || |  |   | |||
Db    1409 ACTTAGAGAACAGCCCACTGAAAACGGAACTGGAGGCAAATGGTAGATCACTCGAAAAGG 1468

Qy     621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
           || |  ||||    |||    ||  || | | || ||| |||||  |   ||  || ||
Db    1469 TAAATAATTCCGTGAGCAAGACAGGAAGCGTAGACACACTTCATAATAAAGAGGGAACAC 1528

Qy     681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
            |||||  | | || |  ||   | |  |  |  |  |||| ||    |||  | | |
Db    1529 TGGAACAACGAGAACAGAACGAAAATCTGCCAAGTGATAAAAGTGACTCTATGGTAGACA 1588

Qy     741 ag 742
           ||
```

```
Db   1589 AG 1590


RESULT 13
US-08-521-872-13
; Sequence 13, Application US/08521872
; Patent No. 6015682
; GENERAL INFORMATION:
;   APPLICANT:  Young, Richard A.
;   APPLICANT:  Koleske, Anthony J.
;   APPLICANT:  Thompson, Craig M.
;   APPLICANT:  Chao, David M.
;   TITLE OF INVENTION:  No. 6015682el Factors Which Modify Gene
;   TITLE OF INVENTION:  Transcription and Methods of Use Therefor
;   NUMBER OF SEQUENCES:  37
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET:  Two Militia Drive
;     CITY:  Lexington
;     STATE:  Massachusetts
;     COUNTRY:  USA
;     ZIP:  02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/521,872
;     FILING DATE:  31-AUG-1995
;     CLASSIFICATION:  436
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/218,265
;     FILING DATE:  25-MAR-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Granahan, Patricia
;     REGISTRATION NUMBER:  32,227
;     REFERENCE/DOCKET NUMBER:  WHI94-03A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  617-861-6240
;     TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  4849 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  148..4407
;   PUBLICATION INFORMATION:
;     DOCUMENT NUMBER:  US 08/218,265
;     FILING DATE:  25-MAR-1994
US-08-521-872-13


  Query Match          5.0%;  Score 38.8;  DB 3;  Length 4849;
  Best Local Similarity   47.5%;  Pred. No. 0.27;
  Matches  115;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy     501 aaaagaaggttgaagaaaattcttgggattctgagagtctccgtgagactgtttcacaga 560
           | || ||  |  | || |||   |    |  ||| ||    |     || || ||| |
Db    1349 ATAATAACATCAATGAGAATGAATTATTTAATGATAGGAAACAAACTACAGTATCAAATG 1408

Qy     561 aggatgtgtgtgtacccaaggctacacatcaaaaagaaatggataaaataagtggaaaat 620
           |   | |      ||   |   ||  | |   | | ||  | || |  |   | |||
Db    1409 ACTTAGAGAACAGCCCACTGAAAACGGAACTGGAGGCAAATGGTAGATCACTCGAAAAGG 1468

Qy     621 tagaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaa 680
           || |  ||||    |||    ||  || | | || ||| |||||  |   ||  || ||
Db    1469 TAAATAATTCCGTGAGCAAGACAGGAAGCGTAGACACACTTCATAATAAAGAGGGAACAC 1528

Qy     681 gggaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaaga 740
           |||||  | | || |  ||    | |  |  |  |  |  ||||  ||    |||  | | |
Db    1529 TGGAACAACGAGAACAGAACGAAAATCTGCCAAGTGATAAAAGTGACTCTATGGTAGACA 1588

Qy     741 ag 742
           ||
Db    1589 AG 1590


RESULT 14
US-09-182-816-22
; Sequence 22, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
;  APPLICANT: Wisnewski, Nancy
;  APPLICANT: Silver, Gary M.
;  APPLICANT: Lo, Katherine C.
;  APPLICANT: Brandt, Kevin S.
;  TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
;  TITLE OF INVENTION: PROTEINS AND USES THEREOF
;  FILE REFERENCE: FC-3-C1
;  CURRENT APPLICATION NUMBER: US/09/182,816
;  CURRENT FILING DATE: 1998-10-29
;  EARLIER APPLICATION NUMBER: 08/989,510
;  EARLIER FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1736
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (159)..(1553)
US-09-182-816-22


  Query Match          4.9%;  Score 38.2;  DB 4;  Length 1736;
  Best Local Similarity   54.7%;  Pred. No. 0.27;
  Matches   76;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     603 ataaaataagtggaaaattagaagattcaactagcctatcaaaaatcttggatacagttc 662
           |||| ||||    |||| || | |     |||     || ||  || | |  |  | |||
Db    1583 ataatataatgttaaaaataaatgtaattactgtgaaataaacgatatggattttatttc 1642

Qy     663 attcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaacgtacaggaaaaa 722
           |  ||||| ||| |  ||   ||    ||||| |  |     || ||   | |  |||||
Db    1643 aaacttgtcaaatataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1702

Qy     723 tggaacaaatgaaaaagaa 741
              || |||  ||||| ||
Db    1703 aaaaaaaaaaaaaaaaaaa 1721


RESULT 15
US-09-182-816-24/c
; Sequence 24, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
;  APPLICANT: Wisnewski, Nancy
;  APPLICANT: Silver, Gary M.
;  APPLICANT: Lo, Katherine C.
;  APPLICANT: Brandt, Kevin S.
;  TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
;  TITLE OF INVENTION: PROTEINS AND USES THEREOF
;  FILE REFERENCE: FC-3-C1
;  CURRENT APPLICATION NUMBER: US/09/182,816
;  CURRENT FILING DATE: 1998-10-29
;  EARLIER APPLICATION NUMBER: 08/989,510
;  EARLIER FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
```

; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-182-816-24

Query Match 4.9%; Score 38.2; DB 4; Length 1736;
Best Local Similarity 54.7%; Pred. No. 0.27;
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy     603 ataaaataagtggaaaattagaagattcaactagcctatcaaaaatcttggatacagttc 662
           |||| ||||    |||| || | |     |||     || ||  || | |  |  | |||
Db     154 ATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAACGATATGGATTTTATTTC 95

Qy     663 attcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaacgtacaggaaaaa 722
           |  ||||| ||| |  ||   ||    ||||| |  |     || ||   | |  |||||
Db      94 AAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 35

Qy     723 tggaacaaatgaaaaagaa 741
              || |||  ||||| ||
Db      34 AAAAAAAAAAAAAAAAAA 16
```

Search completed: May 1, 2001, 00:33:45
Job time: 56569 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2001, 12:17:47 ; Search time 12176.5 Seconds
(without alignments)
9.351 Million cell updates/sec

Title: US-09-721-183-5
Perfect score: 779
Sequence: 1 gtatacattctttattaatc.....gaaactgtcagaagcaaaaa 779

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 73081774 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	503	64.6	574	146	BF328582	BF328582 RC5-BN019
c	2	467.6	60.0	490	104	AI951118	AI951118 wx63g05.x
	3	266.8	34.2	443	168	BF746270	BF746270 RC1-BT025
	4	266.8	34.2	443	168	BF746340	BF746340 RC1-BT025
	5	127.4	16.4	431	162	BE062233	BE062233 RC1-BT025
c	6	111.6	14.3	540	107	AU146672	AU146672 AU146672
	7	101.8	13.1	344	223	AZ576125	AZ576125 AST-T32E0
c	8	95.2	12.2	491	15	AI084496	AI084496 ox78g05.x
	9	94.2	12.1	291	5	AA331953	AA331953 EST35833
c	10	93.2	12.0	427	104	AI989660	AI989660 ws26h05.x
c	11	93.2	12.0	460	102	AI825717	AI825717 wb75g06.x
c	12	93.2	12.0	611	174	BG150291	BG150291 7j98e03.x
	13	87.4	11.2	546	141	BE929777	BE929777 RC3-GN004
c	14	87.2	11.2	543	219	AZ301121	AZ301121 UP_338-19
	15	84.8	10.9	478	170	BF894500	BF894500 RC1-MT013
	16	84.8	10.9	521	170	BF895056	BF895056 RC1-MT013
	17	81.8	10.5	484	168	BF747969	BF747969 RC1-BN041
	18	81.6	10.5	432	170	BF894503	BF894503 RC1-MT013

	19	81.6	10.5	460	170	BF894504	BF894504 RC1-MT013
	20	81.6	10.5	485	170	BF895048	BF895048 RC1-MT013
c	21	81.6	10.5	510	110	AW090234	AW090234 xc92h10.x
	22	81.6	10.5	549	170	BF895055	BF895055 RC1-MT013
	23	81.2	10.4	165	232	HSMC18H11	X88328 H.sapiens D
	24	80.8	10.4	254	160	BB580017	BB580017 BB580017
	25	80	10.3	434	170	BF894486	BF894486 RC1-MT013
	26	80	10.3	452	170	BF894492	BF894492 RC1-MT013
c	27	80	10.3	695	229	AG010124	AG010124 Homo sapi
c	28	78.6	10.1	727	202	AQ115695	AQ115695 RPCI11-49
c	29	77.6	10.0	358	155	N90591	N90591 zb42e02.s1
c	30	77.6	10.0	365	2	AA088258	AA088258 zl82b06.s
c	31	77.6	10.0	408	20	AI440090	AI440090 ti56f05.x
c	32	77.6	10.0	460	11	AA781851	AA781851 ai61g11.s
c	33	77.6	10.0	467	18	AI281642	AI281642 qu48b06.x
c	34	77.6	10.0	517	18	AI281643	AI281643 qu48b07.x
c	35	77.6	10.0	568	23	AI683185	AI683185 tw49g04.x
c	36	77.6	10.0	915	17	AI241666	AI241666 qu70c08.x
	37	76.6	9.8	608	162	BE090060	BE090060 RC6-BT071
c	38	76	9.8	363	4	AA229820	AA229820 nc51a03.r
	39	75.4	9.7	430	170	BF894484	BF894484 RC1-MT013
	40	75	9.6	532	227	B17512	B17512 345O2.TPB C
c	41	74.4	9.6	700	229	AG011606	AG011606 Homo sapi
c	42	74.4	9.6	710	229	AG011605	AG011605 Homo sapi
	43	74.2	9.5	426	140	BE814995	BE814995 MR2-BN011
	44	73.6	9.4	477	209	AQ667184	AQ667184 HS_2109_A
	45	73.4	9.4	665	227	B15703	B15703 345O2.TP CI

ALIGNMENTS

RESULT 1
BF328582
LOCUS BF328582 574 bp mRNA EST 22-NOV-2000
DEFINITION RC5-BN0192-010900-025-E09 BN0192 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF328582
VERSION BF328582.1 GI:11299317
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 574)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC5&t2=RC5-BN0192-010900-025-E09&t3=2000-09-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 525.
FEATURES Location/Qualifiers
source 1. .574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BN0192"
/dev_stage="Adult"
/note="Organ: breast_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 200 a 104 c 130 g 140 t
ORIGIN

Query Match 64.6%; Score 503; DB 146; Length 574;
Best Local Similarity 97.0%; Pred. No. 3.9e-112;
Matches 523; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

```
Qy    106 attgagagcagatgagatactcccatcagaatccaaacaaaaggactatgaagaaagttc 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     36 ATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATGAAGAAAGTTC 95

Qy    166 ttgggattc-tgagagtctctgtgagactgtttcacagaaggatgtgtgtttacccaagg 224
          |||||   | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 TTGGGGGGCGTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTTTACCCAAGG 155

Qy    225 ctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctcctgttaaag 284
          |||| | ||||||||||||||||||||||||||||||||||| |||||||| ||| |
Db    156 CTACGCGTCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGAGTCTCCTGATAATG 215

Qy    285 atggtcttctgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaat 344
          ||||| |||||||||||  |||| ||||||||||||||||||||||||||||||||||||
Db    216 ATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAAT 275

Qy    345 tgatggacatgcaaactttcaaagcagagcctcccgagaagccatctgccttcgagcctg 404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 TGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCTTCGAGCCTG 335

Qy    405 ccattgaaatgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacat 464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 CCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACAT 395

Qy    465 tgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattctt 524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 TGAGAGCAGATCAGATGTTCCCTTCAGAATCAAAACAAAAGAAGGTTGAAGAAAATTCTT 455

Qy    525 gggattctgagagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggcta 584
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 GTGATTCTGAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTGTACCCAAGGCTA 515

Qy    585 cacatcaaaaagaaatggataaaataagtggaaaattagaagattcaactagcctatca 643
          |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||
Db    516 CACATCAAAAAGATATGGATAAAATAAGTGGAAAATTAGAAGATTCAACTGGCCTATCA 574
```

RESULT 2
AI951118/c
LOCUS AI951118 490 bp mRNA EST 09-MAR-2000
DEFINITION wx63g05.x1 NCI_CGAP_Br18 Homo sapiens cDNA clone IMAGE:2548376 3', mRNA sequence.
ACCESSION AI951118
VERSION AI951118.1 GI:5743428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 490)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

```
           Email: Robert_Strausberg@nih.gov
           Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
           Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
           Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
           I.M.A.G.E. Consortium DNA Sequencing by: Washington University
           Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 2157   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2548376"
                     /clone_lib="NCI_CGAP_Br18"
                     /tissue_type="four pooled high-grade tumors, including two
                     primary tumors and two metastatic to ovary"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies."
BASE COUNT      116 a      98 c      89 g     187 t
ORIGIN


  Query Match             60.0%;  Score 467.6;  DB 104;  Length 490;
  Best Local Similarity   97.1%;  Pred. No. 1.5e-103;
  Matches  476;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    232 tcaaaaagaaatagataaaataaatggaaaattagaagggtctcctgttaaagatggtct 291
          |||||||||||||||||||||||||||||||||||||||| |||||||| ||| |||||| |
Db    490 TCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGAGTCTCCTGATAATGATGGTTT 431

Qy    292 tctgaaggctaactgcggaatgaaagtttctattccaactaaagccttagaattgatgga 351
          |||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||
Db    430 TCTGAAGGCTCCCTGCAGAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAATTGATGGA 371

Qy    352 catgcaaactttcaaagcagagcctcccgagaagccatctgccttcgagcctgccattga 411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 CATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCATCTGCCTTCGAGCCTGCCATTGA 311

Qy    412 aatgcaaaagtctgttccaaataaagccttggaattgaagaatgaacaaacattgagagc 471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGAATGAACAAACATTGAGAGC 251

Qy    472 agatcagatgttcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattc 531
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    250 AGATCAGATGTTCCCTTCAGAATCAAAACAAAAGAACGTTGAAGAAAATTCTTGGGATTC 191

Qy    532 tgagagtctccgtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatca 591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 TGAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTGTACCCAAGGCTACACATCA 131

Qy    592 aaaagaaatggataaaataagtggaaaattagaagattcaactagcctatcaaaaatctt 651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 AAAAGAAATGGATAAAATAAGTGGAAAATTAGAAGATTCAACTAGCCTATCAAAAATCTT 71

Qy    652 ggatacagttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaacg 711
          |||||||||||||||||||||| |||||||||||||||||| | |||||| | |||
Db     70 GGATACAGTTCATTCTTGTGAAAGAACAAGGGAACTTCAAAAAGACCCCTGTGACCCACG 11

Qy    712 tacaggaaaa 721
          | ||||||||
Db     10 TTCAGGAAAA 1


RESULT  3
BF746270
LOCUS       BF746270      443 bp    mRNA            EST       10-JAN-2001
DEFINITION  RC1-BT0254-071100-119-a03 BT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF746270
VERSION     BF746270.1  GI:12072946
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-BT0254-
            071100-119-a03&t3=2000-11-07&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0254"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      150 a      84 c     103 g     106 t
ORIGIN


  Query Match             34.2%;  Score 266.8;  DB 168;  Length 443;
  Best Local Similarity   95.8%;  Pred. No. 9.7e-55;
  Matches  274;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy     36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
          || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    158 CCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 217

Qy     96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    218 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 277

Qy    156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    278 AAGAAAGTTCTGGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 337

Qy    216 tacccaaggctacacatcaaaaagaaatagatagaaataaatggaaaattagaagggtctc 275
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    338 TACCCAAGTCTACACATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGAGTCTC 397

Qy    276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttc 321
          ||| ||| |||||| |||||||||||  |||| ||||||||||||
Db    398 CTGATAATGATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTTC 443
```

```
RESULT   4
BF746340
LOCUS       BF746340      443 bp    mRNA            EST       10-JAN-2001
DEFINITION  RC1-BT0254-081100-119-a03 BT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF746340
VERSION     BF746340.1  GI:12073016
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-BT0254-
            081100-119-a03&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0254"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      150 a      84 c     103 g     106 t
ORIGIN


  Query Match           34.2%;  Score 266.8;  DB 168;  Length 443;
  Best Local Similarity   95.8%;  Pred. No. 9.7e-55;
  Matches  274;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      36 ccatttagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaaga 95
           || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158 CCTTCGAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGA 217

Qy      96 atgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactatg 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     218 ATGAACAAACATTGAGAGCAGATGAGATACTCCCATCAGAATCCAAACAAAAGGACTATG 277

Qy     156 aagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgtgtgtt 215
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     278 AAGAAAGTTCTGGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTGTT 337

Qy     216 tacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaagggtctc 275
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     338 TACCCAAGTCTACACATCAAAAAGAAATAGATAAAATAAATGGAAAATTAGAAGAGTCTC 397

Qy     276 ctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttc 321
           ||| ||| |||||| |||||||||||  |||| ||||||||||||
Db     398 CTGATAATGATGGTTTTCTGAAGGCTCCCTGCAGAATGAAAGTTTC 443


RESULT   5
BE062233
LOCUS       BE062233      431 bp    mRNA            EST       09-JUN-2000
DEFINITION  RC1-BT0254-050400-110-c11 BT0254 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE062233
VERSION     BE062233.1  GI:8406883
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC1-BT0254-050
            400-110-c11&t3=2000-04-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 284
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0254"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      121 a      95 c      97 g     118 t
ORIGIN


  Query Match           16.4%;  Score 127.4;  DB 162;  Length 431;
  Best Local Similarity   87.6%;  Pred. No. 7.4e-21;
  Matches  162;  Conservative    0;  Mismatches   21;  Indels    2;  Gaps    2;

Qy     124 actcccatcagaatccaaacaaaaggactatgaagaaagttcttgggattctgagagtct 183
           | || |||    || |  | ||   ||||||||||||||||||||||||||||||||||
Db     247 ATTCTCATGGTCATACCTATAACCTGACTATGAAGAAAGTTCTTGGGATTCTGAGAGTCT 306

Qy     184 ctgtgag-actgtttcacagaaggatgtgtgtttacccaaggctacacatcaaaaagaaa 242
           ||||||| |||||||||||||||||||||||| ||||| ||||||||||||||||||
Db     307 CTGTGAGAACTGTTTCACAGAAGGATGTGTGTTTA-CCAAGTCTACACATCAAAAAGAAA 365
```

```
Qy    243 tagataaaataaatggaaaattagaagggtctcctgttaaagatggtcttctgaaggcta 302
          ||||||||||||||||||| |||||||| |||||||| ||| |||||| ||||||||||
Db    366 TAGATAAAATAAATGGAACATTAGAAGAGTCTCCTGATAATGATGGTTTTCTGAAGGCTC 425

Qy    303 actgc 307
          ||||
Db    426 CCTGC 430


RESULT   6
AU146672/c
LOCUS       AU146672      540 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU146672 HEMBB1 Homo sapiens cDNA clone HEMBB1001175 3', mRNA
            sequence.
ACCESSION   AU146672
VERSION     AU146672.1  GI:11008193
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
            ,S. and Isogai,T.).
  TITLE     HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
            Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
            ,Y., Nagai,T., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBB1001175"
                     /clone_lib="HEMBB1"
                     /tissue_type="whole embryo, mainly body"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      126 a    128 c     82 g    200 t      4 others
ORIGIN


  Query Match             14.3%;  Score 111.6;  DB 107;  Length 540;
  Best Local Similarity   57.2%;  Pred. No. 5.4e-17;
  Matches  242;  Conservative    0;  Mismatches  175;  Indels    6;  Gaps    2;

Qy     95 aatgaacaaacattgagagcagatgagatactcccatcagaatccaaacaaaaggactat 154
          | ||  |||| | |||   ||| |||   | ||| || ||     ||| |||| ||
Db    454 AGTGCCCAAAATTATAGAAGTGATATGATGTCCGCATTAGGATTAGGNCAAGAGGAAGAT 395

Qy    155 gaagaaagttcttgggattctgagagtctctgtgagactgtttcacagaaggatgt---g 211
            ||||    |||||||||||||||| ||| ||||| | || |||||||| ||||
Db    394 ATAGAATCACCTTGGGATTCTGAGAGTATCTCTGAGAATTTTCCACAGAAGTATGTTGAT 335

Qy    212 tgtttacccaaggctacacatcaaaaagaaatagataaaataaatggaaaattagaaggg 271
          |||| |   |||| || |   |||||||| | ||| |  ||||| | ||  |||||
Db    334 CCTTTAGCTGGGGCTGCAGACGGAAAAGAAAAAAATATAGGAAATGAACAAGCAGAAGAT 275

Qy    272 tctcctgttaaagatggtcttctgaaggctaactgcggaatgaaagtttctattccaact 331
              | || |  |  |    ||| |      |      | ||  |  |||  | |
Db    274 GTGTTTTATATACCTTCTTGCATGAGTGGATCAAGAAACTTTAAGATGGCTAAACTA--- 216
```

```
Qy    332 aaagccttagaattgatggacatgcaaactttcaaagcagagcctcccgagaagccatct 391
          | |   |  |    | | ||||| |     | |    ||| |||| ||||       |
Db    217 GAGGATACAAGAAATGTAGGCATGCCAGTAGCCCACATGGAGTCTCCTGAGAGATATCTT 158

Qy    392 gccttcgagcctgccattgaaatgcaaaagtctgttccaaataaagccttggaattgaag 451
          |||  ||||| |||||||||| || | ||||||||||||||||    | | |||||
Db    157 CACTTGAAGCCTACCATTGAAATGAAAGATTCTGTTCCAAATAAAGCAGGAGGAATGAAG 98

Qy    452 aatgaacaaacattgagagcagatcagatgttcccttcagaatcaaaacaaaagaaggtt 511
          ||| ||||||||  | ||||| | |      |     || | ||  | || |  ||
Db     97 GATGTACAAACATCCAAAGCAGCTGAACATGACTTAGAAGTAGCATCAGAAGAAGAGCAA 38

Qy    512 gaa 514
          |||
Db     37 GAA 35


RESULT   7
AZ576125
LOCUS       AZ576125      344 bp    DNA             GSS       06-DEC-2000
DEFINITION  AST-T32E0043 Genetrap T47D Human Breast Carcinoma Library Homo
            sapiens genomic 5', DNA sequence.
ACCESSION   AZ576125
VERSION     AZ576125.1  GI:11562436
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Henkel,G., Liyanage,M., Pratt,E., Huang,D., Riley,M., Bernardino,A.
            , Durick,K. and Pollok,B.
  TITLE     Exon-trap tags from a T47D GenomeScreen(TM) Library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Greg Henkel
            Gene Expression
            Aurora Biosciences Corp.
            11010 Torreyana Road, San Diego, CA 92121, USA
            Tel: 8584048436
            Fax: 8584046719
            Email: henkelg@aurorabio.com
            Pools of cells were isolated from a GenomeScreen(TM) library. The
            library of cells was generated by retroviral integration of a gene
            tagging element consisting of: 1) A promoterless beta-lactamase
            proceeded by a splice acceptor as a reporter for gene expression;
            2) A promoter driving neomycin resistance followed by a splice
            donor to trap downstream exons. 3' RACE from neomycin gene was
            performed using total RNA from isolated pools. Output was shotgun
            cloned in pAmp-1 and used to transform DH5-alpha competent
            bacteria. 5' ends of reported sequences were immediately preceded
            by splice donor from the trapping construct.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Genetrap T47D Human Breast Carcinoma Library"
                     /tissue_type="Carcinoma"
                     /cell_type="Epithelial"
                     /cell_line="T47D"
                     /note="Organ: Breast; Vector: pAmp-1; 3' RACE of total RNA
                     from genetrap pools; shotgun clone in pAmp-1 and  used to
                     transform DH5-alpha competent bacteria."
BASE COUNT      110 a     69 c     79 g     86 t
ORIGIN


  Query Match             13.1%;  Score 101.8;  DB 223;  Length 344;
  Best Local Similarity   84.6%;  Pred. No. 1.2e-14;
  Matches  126;  Conservative    0;  Mismatches   22;  Indels    1;  Gaps    1;

Qy    483 tcccttcagaatcaaaacaaaagaaggttgaagaaaattcttgggattctgagagtctcc 542
```

```
          |||| |||||||| ||||||||| | | |||||||||||||||||||| ||||||| |||
Db    139 TCCCATCAGAATCCAAACAAAAGGATGATGAAGAAAATTCTTGGGATTTTGAGAGTTTCC 198

Qy    543 gtgagactgtttcacagaaggatgtgtgtgtacccaaggctacacatcaaaaagaaatgg 602
          |||||||  || |||||| ||||||||| |||||||||||||||||||||||||||| | |
Db    199 TTGAGACTCCTT-ACAGAATGATGTGTGTTTACCCAAGGCTACACATCAAAAAGAATTCG 257

Qy    603 ataaaataagtggaaaattagaagattca 631
          |||   |||||||||||||  |   ||||
Db    258 ATACCTTAAGTGGAAAATTGAAGATTTCA 286


RESULT   8
AI084496/c
LOCUS       AI084496      491 bp    mRNA            EST       24-SEP-1998
DEFINITION  ox78g05.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1662488
            3' similar to TR:Q15694 Q15694 PROTEIN IMMUNO-REACTIVE WITH
            ANTI-PTH POLYCLONAL ANTIBODIES ;, mRNA sequence.
ACCESSION   AI084496
VERSION     AI084496.1  GI:3422919
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2073   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1662488"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      125 a    104 c     74 g    188 t
ORIGIN


  Query Match           12.2%;  Score 95.2;  DB 15;  Length 491;
  Best Local Similarity  70.6%;  Pred. No. 5.1e-13;
  Matches  127;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    599 atggataaaataagtggaaaattagaagattcaactagcctatcaaaaatcttggataca 658
          || ||  || |  |     | |  |||||||  ||||||||  ||||||  |||| ||
Db    256 ATTGAACAACTTGGAATGGAGTGTAAAGATTCTGTTAGCCTATTGAAAATCCAGGATGCA 197

Qy    659 gttcattcttgtgaaagagcaagggaacttcaaaaagatcactgtgaacaacgtacagga 718
          | || ||| |||||||||  |   |||||| ||||| |||||||||||| || ||||| |
Db    196 GCTCTTTCATGTGAAAGATTATTAGAACTTAAAAAAAATCACTGTGAACTACTTACAGTA 137

Qy    719 aaaatggaacaaatgaaaaagaagttttgtgtactgaaaaagaaactgtcagaagcaaaa 778
          |||||  || ||||| || | ||| ||  ||||||  ||| | | || || ||| |||||
Db    136 AAAATTAAAAAAATGGAAGACAAGGTTAATGTACTACAAAGGGAGCTATCTGAAACAAAA 77


RESULT   9
AA331953
LOCUS       AA331953      291 bp    mRNA            EST       21-APR-1997
DEFINITION  EST35833 Embryo, 8 week I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA331953
VERSION     AA331953.1  GI:1984195
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):133580"
                     /db_xref="taxon:9606"
                     /clone_lib="Embryo, 8 week I"
                     /dev_stage="embryo, 8 wks"
                     /note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      121 a     38 c     57 g     74 t      1 others
ORIGIN


  Query Match           12.1%;  Score 94.2;  DB 5;  Length 291;
  Best Local Similarity  75.5%;  Pred. No. 8.1e-13;
  Matches  117;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy    624 aagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaaggg 683
          |||||||  ||||||||  ||||||  |||| ||| || ||| |||||||||  |  |
Db      7 AAGATTCTGTTAGCCTATTGAAAATCCAGGATGCAGCTCTTTCATGTGAAAGATTATTAG 66

Qy    684 aacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagt 743
          ||||| ||||| |||||||||||| || ||||| ||||||  || ||||| || | |||
```

```
Db     67 AACTTAAAAAAAATCACTGTGAACTACTTACAGTAAAAATTAAAAAAATGGAAGACAAGG 126

Qy    744 tttgtgtactgaaaaagaaactgtcagaagcaaaa 778
          ||  ||||||  ||| | | || || ||| |||||
Db    127 TTAATGTACTACAAAGGGAGCTATCTGAAACAAAA 161


RESULT 10
AI989660/c
LOCUS       AI989660      427 bp    mRNA            EST       08-SEP-1999
DEFINITION  ws26h05.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2498361 3'
            similar to contains element MER30 repetitive element ;, mRNA
            sequence.
ACCESSION   AI989660
VERSION     AI989660.1  GI:5836541
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2498361"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT       86 a     92 c     73 g    176 t
ORIGIN


  Query Match             12.0%;  Score 93.2;  DB 104;  Length 427;
  Best Local Similarity   88.6%;  Pred. No. 1.5e-12;
  Matches  101;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    625 agattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaaggga 684
          |||||||||||||||||||| | |||||| | |||||| ||||  ||||||||||||||||
Db    115 AGATTCAACTAGCCTATCAATATTCTTGGGTGCAGTTCCTTCTCATGAAAGAGCAAGGGA 56

Qy    685 acttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaa 738
          |||| ||||| ||||||||||||||| ||||| ||||||  ||||||||||||
Db     55 ACTTAAAAAATATCACTGTGAACAACTTACAGCAAAAATAAAACAAATGAAAAA 2


RESULT 11
AI825717/c
LOCUS       AI825717      460 bp    mRNA            EST       17-DEC-1999
DEFINITION  wb75g06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2311546 3',
            mRNA sequence.
ACCESSION   AI825717
VERSION     AI825717.1  GI:5446388
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 549   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2311546"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT       91 a     99 c     79 g    191 t
ORIGIN


  Query Match             12.0%;  Score 93.2;  DB 102;  Length 460;
  Best Local Similarity   88.6%;  Pred. No. 1.5e-12;
  Matches  101;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    625 agattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaaggga 684
          |||||||||||||||||||| | |||||| | |||||| ||||  ||||||||||||||||
Db    115 AGATTCAACTAGCCTATCAATATTCTTGGGTGCAGTTCCTTCTCATGAAAGAGCAAGGGA 56

Qy    685 acttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaa 738
          |||| ||||| ||||||||||||||| ||||| ||||||  ||||||||||||
Db     55 ACTTAAAAAATATCACTGTGAACAACTTACAGCAAAAATAAAACAAATGAAAAA 2


RESULT 12
BG150291/c
LOCUS       BG150291      611 bp    mRNA            EST       05-FEB-2001
DEFINITION  7j98e03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3443020 3'
            similar to contains element MER30 repetitive element ;, mRNA
            sequence.
ACCESSION   BG150291
```

```
VERSION     BG150291.1  GI:12662321
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3443020"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      122 a    131 c    102 g    256 t
ORIGIN
```

```
  Query Match             12.0%;  Score 93.2;  DB 174;  Length 611;
  Best Local Similarity   88.6%;  Pred. No. 1.6e-12;
  Matches  101;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    625 agattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggga 684
          |||||||||||||||||||| | |||||| | |||||| ||||  ||||||||||||||||
Db    115 AGATTCAACTAGCCTATCAATATTCTTGGGTGCAGTTCCTTCTCATGAAAGAGCAAGGGA 56

Qy    685 acttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaa 738
          |||| ||||| ||||||||||||||| ||||| ||||||  ||||||||||||
Db     55 ACTTAAAAAATATCACTGTGAACAACTTACAGCAAAAATAAAACAAATGAAAAA 2
```

```
RESULT  13
BE929777
LOCUS       BE929777      546 bp    mRNA            EST       02-OCT-2000
DEFINITION  RC3-GN0042-170800-022-d11 GN0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE929777
VERSION     BE929777.1  GI:10455853
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC3-GN0042-170
            800-022-d11&t3=2000-08-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 546.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0042"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      224 a     70 c    119 g    133 t
ORIGIN
```

```
  Query Match             11.2%;  Score 87.4;  DB 141;  Length 546;
  Best Local Similarity   75.2%;  Pred. No. 4.1e-11;
  Matches  109;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy    634 tagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaacttcaaaa 693
          ||||||||  ||||||  |||| ||| || ||| |||||||||  | | |||||| ||||
Db     56 TAGCCTATGTAAAATCCAGGATGCAGCTCTTTCATGTGAAAGATTATGAGAACTTAAAAA 115

Qy    694 agatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagttttgtgtact 753
          | |||||||||| | || || || |||||| || ||||| || | ||| ||  ||||||
Db    116 AAATCACTGTGAGCTACTTAGAGTAAAAATTAAAAAAATGGAAGACAAGGTTAATGTACT 175

Qy    754 gaaaaagaaactgtcagaagcaaaa 778
          ||| | | || || ||| |||||
Db    176 ACAAAGGGAGCTATCTGAAACAAAA 200
```

```
RESULT  14
AZ301121/c
LOCUS       AZ301121      543 bp    DNA             GSS       29-DEC-2000
DEFINITION  UP_338-19G_T7 RPCI11 Human Male BAC Library Homo sapiens genomic
            clone RP11-338G19, DNA sequence.
ACCESSION   AZ301121
VERSION     AZ301121.2  GI:11995798
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Cheung,V.G., Dalrymple,H.L., Narasimhan,S., Watts,J., Schuler,G.,
            Raap,A.K., Morley,M. and Bruzel,A.
  TITLE     A resource of mapped human bacterial artificial chromosome clones
```

```
  JOURNAL   Genome Res. 9 (10), 989-993 (1999)
  MEDLINE   99455100
COMMENT     On Dec 29, 2000 this sequence version replaced gi:9795716.
            Contact: Arcaro MA, Morley M, Burdick J, Cheung VG
            Department of Pediatrics
            University of Pennsylvania
            3516 Civic Center Blvd, ARC 516, Philadelphia, PA 19104, USA
            Tel: 215 590 2664
            Fax: 215 590 3709
            Email: ARCARO@email.chop.edu
            Plate: 338  row: G  column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-338G19"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      171 a    107 c     83 g    182 t
ORIGIN


  Query Match             11.2%;  Score 87.2;  DB 219;  Length 543;
  Best Local Similarity   72.4%;  Pred. No. 4.6e-11;
  Matches  113;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     623 gaagattcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagg 682
           | ||||||   |||||||||| |||||    ||| |||||| ||| || |||     |
Db     423 GTAGATTCTGTTAGCCTATCGAAAATGCAAGATGCAGTTCTTTCATGGGAACACTTATTA 364

Qy     683 gaacttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaag 742
           |||||| | ||  |||| |||||||||| ||||| ||| ||  |||||||| |||| | |
Db     363 GAACTTAAGAATAATCATTGTGAACAACTTACAGTAAATATTAAACAAATGGAAAATATG 304

Qy     743 ttttgtgtactgaaaaagaaactgtcagaagcaaaa 778
           || |||||||  ||||| | || |  |||||||||
Db     303 GTTGGTGTACTACAAAAGGAGCTATGTGAAGCAAAA 268


RESULT  15
BF894500
LOCUS       BF894500      478 bp    mRNA            EST       18-JAN-2001
DEFINITION  RC1-MT0135-171100-021-f02 MT0135 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF894500
VERSION     BF894500.1  GI:12285959
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-MT0135-
            171100-021-f02&t3=2000-11-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 45
            High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0135"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      194 a     75 c    109 g    100 t
ORIGIN


  Query Match             10.9%;  Score 84.8;  DB 170;  Length 478;
  Best Local Similarity   72.4%;  Pred. No. 1.7e-10;
  Matches  110;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     627 attcaactagcctatcaaaaatcttggatacagttcattcttgtgaaagagcaagggaac 686
           ||||   ||||||||| |||||    ||| |||||| ||| || |||     |   ||||
Db      94 ATTCTGTTAGCCTATCGAAAATGCAAGATGCAGTTCTTTCATGGGAACACTTATTAGAAC 153

Qy     687 ttcaaaaagatcactgtgaacaacgtacaggaaaaatggaacaaatgaaaaagaagtttt 746
           || | ||  |||| |||||||||| ||||| ||| ||  |||||||| |||| | | ||
Db     154 TTAAGAATAATCATTGTGAACAACTTACAGTAAATATTAAACAAATGGAAAATATGGTTG 213

Qy     747 gtgtactgaaaaagaaactgtcagaagcaaaa 778
           |||||||  ||||| | || |  |||||||||
Db     214 GTGTACTACAAAAGGAGCTATGTGAAGCAAAA 245


Search completed: April 30, 2001, 12:17:51
Job time: 12540 sec
```

This Page Blank (uspto)

This Page Blank (uspto)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☑ OTHER: duplex copies are on opposite ends

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)